Anodot

THE BUILD OR BUY DILEMMA

HOW TO GET THE MOST FROM AN ANOMALY DETECTION SYSTEM

May 2017

THE BUILD OR BUY DILEMMA: HOW TO GET THE MOST FROM AN ANOMALY DETECTION SYSTEM

Leveraging the vast amount of business data available today to better meet customer needs and detect business incidents presents organizations with the challenge of whether to build their own anomaly detection system or buy one ready-made. Before organizations make this critical decision, it is important to weigh the benefits and challenges of each approach.

HIGHLIGHTS & KEY FINDINGS

- The variables affecting business operations and the amount of related data available to analyze is growing at exponential rates, with some studies estimating global data volume will reach 44 zettabytes or 44 trillion gigabytes by 2020.

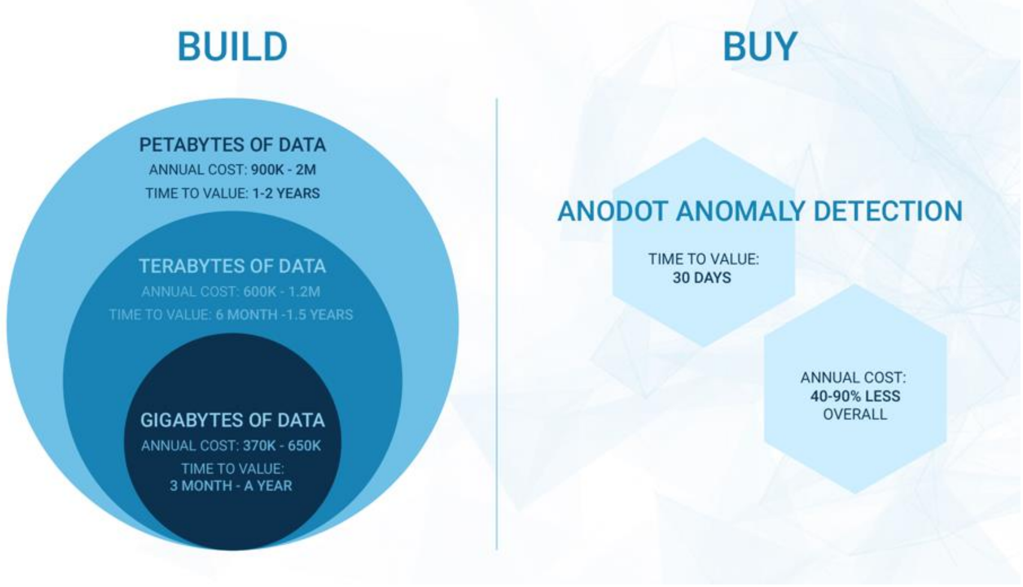

Figure 1: Choose your circle based on how much data you have. Overall, buying an off-the-shelf anomaly detection solution will cost 40-90% less than building, and will bring value in 30 days, as opposed to several months to several years.

- For organizations considering building their own anomaly detection system, the main decision factors at the end of the day are cost, time to value, complexity, and how crucial the system is to the core business of the company.
- Before building or buying an anomaly detection solution, organizations need to consider several key issues, including data volume, expansion/growth plans, budget, how quickly the system is needed, and willingness to invest in data science and other staff.

- Build options are usually only viable for companies with small amounts of uncomplicated data or extremely large, innovative companies that have a dedicated anomaly detection team.
- The cost of the Anodot anomaly detection system is often 40 percent to 90 percent less than the total cost for a company to build its own anomaly detection system, including hiring staff, purchasing software/infrastructure, and maintaining and updating the system.
- Anodot's off-the-shelf anomaly detection solution provides significantly faster time to value than building a solution in-house.
- Anodot's off-the-shelf anomaly detection system provides a more sophisticated, comprehensive solution than the majority of home-grown solutions.

INTRODUCTION

With the volume of global data estimated to grow to 44 zettabytes or 44 trillion gigabytes by 2020,[1] organizations have the ability to leverage information in ways never before seen. Today companies can capture, analyze, slice and dice information to provide meaningful insights into markets and business operations. E-commerce companies can identify faulty checkout flows on certain devices and hot products for specific user segments. Manufacturers can pinpoint production systems running at less than peak efficiency and requiring maintenance. And, advertising technology (ad-tech) companies can discover almost instantly the source of a decline in conversions or ad displays.

21ST CENTURY ANOMALY DETECTION SYSTEMS

As data gets bigger and more complex, organizations increasingly find it challenging to analyze the vast and variable quantity of information using only existing staff and the conventional methods available such as traditional business intelligence (BI) dashboards and alerts. Fortunately, anomaly detection systems (ADS), such as the type of solution offered by Anodot, have entered a truly astounding phase in their ability to provide organizations with information on real-time and future trends. In detecting blips and determining their root cause, what used to take staff days or even weeks can now be accomplished in a matter of minutes using advanced data analytics. But to get the benefits of modern data science and online machine learning, organizations first need to determine whether to build or buy an anomaly detection system.

> **Today's businesses need anomaly detection to ensure cost savings, speedier and improved decision making, and the ability to successfully create new markets, new products, and new services.**

[1] Vernon Turner, "Executive Summary: Data Growth, Business Opportunities, and the IT Imperatives," The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things, an iView published by IDC, April 2014, http://idc.idcimpshowcase.com/showcase/detail.cfm?id=73

THE PATH TO BUILD OR BUY

Whether it's a corporate mandate to stop revenue loss or a broader plan to expand sales, anomaly detection systems can contribute significantly to achieving corporate goals and objectives. But the decision to integrate anomaly detection into an existing corporate system requires a thorough understanding of an organization's budget, capacity, capabilities, and the types and amount of data to be captured. Consider the following:

- ***Size of the company & volume of data.*** Is your company relatively small with only one or two data metrics or key performance indicators (KPIs), or do you have thousands, hundreds of thousands, or millions of metrics?[2]
- ***Budget.*** How much have you budgeted for the system? If you intend to build your own system, have you budgeted for infrastructure, recruiting and hiring data science and other staff, expansion as the company grows, and potentially expensive ongoing system maintenance?
- ***Time to Value.*** How quickly do you need the system? If you intend to build a system, can you afford a loss of revenue from business incidents until you have the anomaly detection system up and running? Is there a business case for buying a system, instead of building one, in order to free up staff to work on other key revenue-generating or mission-critical projects?
- ***Development & Maintenance.*** Does your company want to hire staff—data scientists, user interface (UI) developers, DevOps, programmers, quality assurance (QA) specialists, and data analysts—to develop your own home-grown anomaly detection system? And, do you want to keep them employed to fully maintain the system once it is developed?
- ***Expansion plans.*** Do you see your company growing in the next year to five years? Do you anticipate that the amount and type of data you plan to capture and analyze will expand by 10, 20, or even 50 percent? Are you prepared to update and issue new releases to a home-grown anomaly detection system as your company grows?

[2] Note: What may appear to be a small number of metrics at first tends to increase when the influence of other factors are examined. For example, consider a company tracking just two business KPIs (e.g. the number of purchases and revenue), across its 50 products, in 10 categories, selling in 50 states, across 8 operating systems. The calculation—2 x 50 x 10 x 50 x 8—equals 400,000 metric streams that need to be tracked and correlated for anomalies.

TO BUILD OR BUY, THAT IS THE QUESTION...

The changing nature of business intelligence means that companies increasingly accept that anomaly detection must be a key component of their current and future operations.[3] Traditional BI dashboards or static thresholds no longer suffice; higher levels of customization and flexibility are needed. Table 1 highlights the business factors and Table 2 the cost factors to consider when weighing the pros and cons of a build scenario compared with a buy option.

Table 1. Key Build/Buy Business Decision Factors

Business Factor	How Build Scenario Affects Business	How Buy Scenario Affects Business
Qualified and Experienced Staff	Organization hires, vets, trains and manages its own data scientists, UI developers, programmers, DevOps, & QA.	Anomaly detection system provider handles all hiring and vetting of its experienced team who are knowledgeable in the relevant fields and available to the organization on an as-needed basis .
Quantity of Data	For smaller data sets, limited to a few KPIs, building your own rudimentary anomaly detection system may make sense.	Hundreds, hundreds of thousands, or millions of KPIs and dimensions can be captured/analyzed using an off-the-shelf anomaly detection solution.
Time to Solution	Extended development time, typically several years for more complex or larger data sets.	Up and running in 30 days or less.
Maintenance	Company maintains its own system, including the creation of new algorithms and programming for additional KPIs and metrics.	Ongoing maintenance is included as part of the ADS service contract.
Ease of Use	Company will need to invest in a UI and advanced algorithms, and still may find itself battling false positives, or business users who cannot find their way around the system.	Commercially available solutions such as Anodot's provide a user-friendly interface, and low false positive rates due to fully validated algorithms, anomaly scoring and other advanced techniques.

[3] Industry examples from businesses that recognize that anomaly detection must be a key component of their current and future operations:

- Dan Frankowski, Pinterest business analytics and data scientist: Learning about your business from anomalous metrics, a blog article published by A Medium Corporation, August 2015; https://medium.com/@Pinterest_Engineering/learning-about-your-business-from-anomalous-metrics-7d2b7ac0bea9
- Unexpected data are good for business, Colloquium on anomaly detection published by the University of Leiden, January 2017, https://www.universiteitleiden.nl/en/news/2017/01/unexpected-data-are-good-for-business
- Towards 99.99% Availability via Intelligent Real-time Alerting, Uber presentation video, published on YouTube, December 2016, https://www.youtube.com/watch?v=_AkX42u72wk&list=PLLEUtp5eGr7Af3AadSb9UiHsH8ZIyAQ7K&index=5

Additional Business Considerations

- Scientists, UI developers, and other technology experts with knowledge of data analytics are highly sought after. Currently there are more unfilled jobs than available qualified staff.
- Data quantities change as companies grow and systems must be adaptable. Key performance indicators such as conversion, page views, and revenue can be tracked on multiple levels, such as by device type, country, and product. Analysis at this level of granularity is complex, and most self-developed solutions will necessarily limit the dimensions or KPIs due to complexity or development time.
- If data changes in complexity or behavior, maintenance can become complicated, particularly if the organization must develop new algorithms, include additional KPIs, or incorporate new systems and networks.
- Depending on the number of variables to be captured in the ADS system, as well as the availability of trained staff to develop the system, some build scenarios can take more than five years to develop, particularly if the company has large, complex, and changing data needs.
- Costs are highly dependent on the type of solution the organization chooses to pursue. For small amounts of data, a rudimentary anomaly detection solution might be adequate. For larger quantities of data metrics, a more robust anomaly detection system that includes scoring and correlation is needed.

Table 2. Key Build/Buy Cost Decision Factors

Data Volume	Costs Types	Build Anomaly Detection	Buy Anomaly Detection
Gigabytes[4]	Software Costs	$20-100k	Overall savings of 40-90% Time to Value: 30 days
	Staff Costs	$350-550k / year (3-5 people) for a rudimentary anomaly detection system	
	Time to Value	90 days to 1.5 years	
	Yearly Maintenance	Additional 2/3 build cost, annually	
Terabytes	Software Costs	~$50-200k	
	Staff Costs	$550-800k / year (5-8 people) for an anomaly detection system without metric correlation	
	Time to Value	6 months – 1.5 years	
	Yearly Maintenance	Additional 2/3 build cost, annually	
Petabytes	Software Costs	~$100-400k	
	Staff Costs	$800k – 2.4m / year (8-18 people)	
	Time to Value	2 – 5 years	
	Yearly Maintenance	Additional 2/3 build cost, annually	

[4] For small companies with small amounts of data, a dashboard and alerts system may be built for an investment of around $5k-$15k in software, plus the salaries of 1-3 people ($60k – 180k annually). The dashboard/alert system will need to be maintained at nearly that level annually after being built. However, these types of systems quickly get complex and expensive depending on the size and type of the data.

Additional Cost Considerations

- Experienced data science staff are in extremely high demand, receive premium salaries, and are difficult to find, recruit, and hire. In addition to data scientists, organizations wishing to build their own solutions in-house will need UI developers, DevOps, programmers, and QA specialists.
- When considering costs, it is important to think about the cost of business incidents that may occur while the organization is building an anomaly detection system, or even after the system is in place. The system must be robust enough to catch every incident. Depending on the type of business and incident, expenses for late detection and time to resolution can cost an additional $50,000 to $200,000 per incident.

OTHER BUILD CONSIDERATIONS

In working with large and small companies in a variety of industries around the globe, Anodot has obtained insight into the types of unique (but not necessarily unusual) challenges an organization can encounter when building their own anomaly detection system. Many large companies such as Facebook and Netflix have publicly shared their own challenging experiences with building their own anomaly detection systems.[5]

Open-source Tools. Some companies feel they have sufficient programming, UI, and DevOps staff to develop their own ADS, basing them on open-source tools. [6] However, organizations need to remember that open-source tools are not simply "install and run" applications. They still require customization in the form of do-it-yourself algorithms and adaptation to ensure the tool targets and captures the correct industry-specific data.

> *"We generally prefer to build all our tools internally, but after working with Anodot, our Chief Data Scientist estimated that it would have taken at least six of our data scientists and engineers more than a year to build something of this caliber, so it was a no-brainer for us to jump on board and take it. Our head of tech ops told us that he's been searching for years for an automated data analysis solution like this."*
>
> *—Rich Galan, Director of Analytics, Rubicon Project*

No Two Data Streams are Alike. The necessary algorithm type will vary depending on the shape of the data stream. In fact, there are dozens of metric distribution varieties, such as smooth, irregular, discrete, sparse, multi-modal, and others, each with its own appropriate algorithms. Developing an anomaly detection solution for smooth metrics is fairly straightforward, but other metric algorithms are more complex.[7]

[5] Examples from corporations that have experienced challenges in developing anomaly detection systems. Facebook post: Top Open Data Problems; Netflix post: RAD Outlier Detection on Big Data

[6] Example of organization which programmed an ADS based on open-source tools: Etsy Skyline

[7] Uber discusses the importance and complexity of selecting the right algorithm for each right data stream in this presentation: Automatic Algorithm Selection for Anomaly Detection: From Prototype to Production.

Alert Fatigue. Accuracy doesn't always equate with useful. An overly sensitive ADS solution can cause even the smallest blip to generate an alert, and the more alerts, the more likely staff will relegate them to an unchecked email folder. Systems need to be designed to score relevant anomalies higher than irrelevant ones, and put multiple anomalies into correlated groups to prevent bombarding users with unnecessary alerts.

Correlation. Anomalies are often related. However, many dashboard and alert systems or home-grown anomaly detection solutions view them independently and will issue one alert for each anomaly, even when the anomalies are connected. A sophisticated anomaly detection solution, such as Anodot, will correlate related alerts so one comprehensive alert (instead of dozens) can point an analysis team in the right direction to identify and solve the problem.

Anomaly Scores. Not all anomalies are equal. Some are more significant than others, and machine learning algorithms need to be able to apply an automated score based on the anomaly's duration and baseline deviation, relative to other anomalies.

Usability. Anomaly detection tools should be user-friendly and accessible at all levels of the corporation—from the data analytics team, to the product marketing staff, to the CEO. Having a well-developed user interface is critical.

THE BUTTERFLY EFFECT: ANOMALY DETECTION SOLUTIONS ARE COMPLEX

As the classic weather model example known as "The Butterfly Effect" explains, an unknown butterfly flapping its wings in one part of the world may create an initial localized and insignificant change in air conditions, yet those changes combine with other changes to build upon each other, eventually significantly impacting future weather conditions globally. Anomaly detection works in a similar way. Time series data, KPIs, metrics, and variables can be more numerous, complex, and interconnected than companies realize initially. For example, a single medium-sized business might have thousands of different variables and sub-variables that influence operations, such as geographies, device types, customer retention and satisfaction, product defects, internet usage and conversion rates, earned value, time to market, etc. A single incident may be the result of thousands or tens of thousands of anomalous variables. An anomaly detection solution must be sophisticated and comprehensive enough to note not only one outcome, but also tell the entire story and point to the root cause.

CONCLUSION

There is a range of potential build options available to an organization interested in adopting anomaly detection as part of its business operations. Being aware of the costs, staffing challenges, and potential pitfalls is critical to ensuring that any home-grown solution not only serves its intended purpose, but also provides a comparable return on investment. For companies with small amounts of data or who do not need real time analytics, a simple home-grown solution may be sufficient. If a company has the relevant data science staff and they can afford to wait until their internal ADS solution is developed, then building their own system might be an option, and we have detailed the steps in building an anomaly detection system in a comprehensive three-part white paper. However in the vast majority of instances, most companies—large and small—find that a comprehensive and sophisticated off-the-shelf anomaly detection system, such as the one offered by Anodot, offers considerably more benefits and faster time to value than those developed in-house.

ABOUT ANODOT

Anodot was founded in 2014, and since its launch in January 2016 has been providing valuable business insights through anomaly detection to its customers in financial technology (fin-tech), ad-tech, web apps, mobile apps, e-commerce and other data-heavy industries. Over 40% of the company's customers are publicly traded companies, including Microsoft, VF Corp, Waze (a Google company), and many others. Anodot's real-time business incident detection uses patented machine learning algorithms to isolate and correlate issues across multiple parameters in real time, supporting rapid business decisions. Learn more at http://www.anodot.com/.