SNOWPLOW
POWERING YOUR DATA JOURNEY

POWERING YOUR DATA JOURNEY: DATA COLLECTION

DATA QUALITY STARTS WITH DATA COLLECTION

Build a data competency center for the business and deliver high quality, meaningful data.

SPEEDREAD: DATA COLLECTION IN 60 SECONDS

> Organizations may not give much thought as to how they collect their behavioral data but this could mean missing out on high-quality, actionable data to drive the use cases they need to succeed.

> By collecting for volume instead of quality, they risk seeding their business with bad data that will undermine decision-making.

> Progressive companies are paying close attention to how they collect their data and are putting solutions in place that enable them to develop a realistic and trustworthy picture of how their customers interact with their products.

> This in turn empowers them to craft market-leading customer experiences, leading to higher engagement, brand loyalty, and improved business performance.

> A data collection strategy starts by identifying the questions you want to answer with data, and then building a suitable data capture strategy.

> Data teams can benefit by embracing data delivery solutions that give them flexibility and control over how data is collected.

> The ability to surface, validate and correct bad data during collection will create confidence across the business that data is high quality, trustworthy, and actionable.

> The head of data is instrumental in creating this data asset and building an intuitive data structure that makes it easy for data consumers to unlock valuable insights.

BEHAVIORAL DATA— THE KEY TO SUCCESS

Access to high-quality behavioral data is what will define successful companies in the future. It will be the key to winning the most valuable customers, keeping them engaged with products and services they will love, and delivering a standard of service they will happily recommend.

Crafting market-leading customer experiences will involve a number of teams relying on data drawn from many different sources, which will be constantly evolving to keep pace with customer demands and competitors. Many of these sources won't be behavioral, such as CRM and advertising data, and people won't be the only users of the data—new technologies such as Machine Learning (ML) and Artificial Intelligence (AI) require high-quality data to deliver their promised efficiencies and intelligence.

As the business uses data in more intensive ways for more complex use cases, it will look to the data team to provide high-quality, actionable data that creates value for everyone in the business. This presents an opportunity for the head of data and the data team to prove their worth—by building a centralized strategic data asset that delivers business success.

Achieving that depends on how well you lay your foundations. And the place to start is with data collection.

It's an issue that typically doesn't garner a lot of attention, but it should—because your data collection processes heavily influence data quality, how much the business can do with its data, and how accurate your analytics are. Paying attention to—and making active decisions about—how data is collected means putting in place the processes and systems that generate, collect, and store high-quality data.

But focusing on the technical issues alone won't deliver a useful data asset.

Understanding the business rationale behind data collection is just as important. How is the data you collect going to help you drive business growth?

Data teams can start by looking at the business's use cases and thinking about what they're trying to achieve. They can then work backwards towards what data they are going to collect and how they define it. In doing so they will build an asset that is reliable, detailed, and easy to work with and understand—in other words, a competency center for data that's impactful across the organization today and in the future.

WHAT IS QUALITY DATA?

Quality data is data the business can rely on when making decisions, knowing (rather than guessing) that it has a complete and accurate picture of what its customers are doing.

Accuracy means that what the data says happened, actually happened.

Completeness means that no part of the picture is missing. You've been able to track what you need without obstruction from adblockers. And you've been able to identify causes of missing data and correct them. Establishing data quality at collection plays an important part in ensuring your business has the high-quality data it needs to thrive. Data collection is often taken for granted, so that the difficulties it presents can be largely underestimated. Mistakes are very easy to make at this stage, so we've outlined below the main traps to avoid.
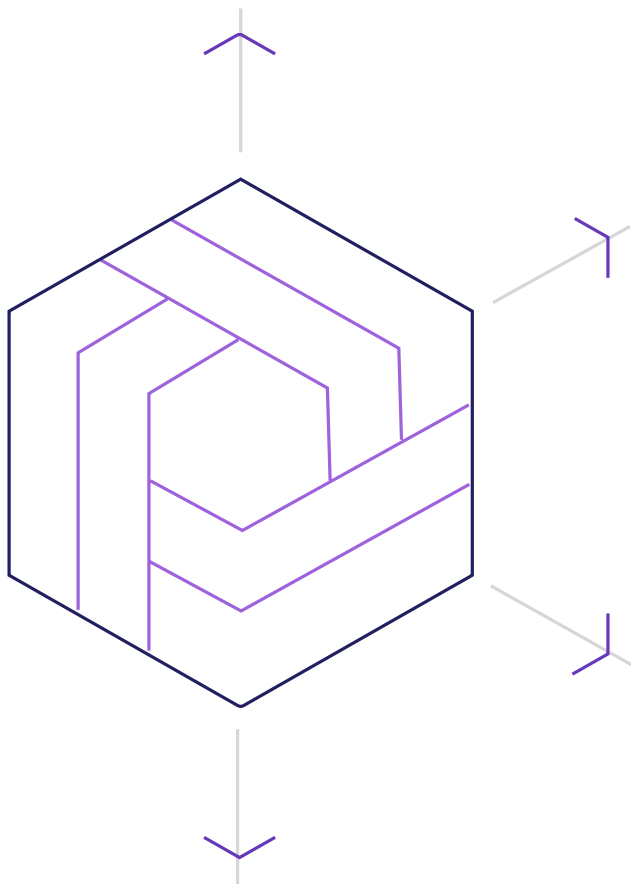
COLLECTING FOR VOLUME RATHER THAN QUALITY

As businesses grow they need to be able to collect and operationalize behavioral data at scale.

Data collection becomes a far more complex endeavor as they track users across an ever-growing number of platforms and develop new use cases as they iterate around evolving customer experiences. Data volumes increase exponentially as the business collects all the behavioral data it can in an effort to understand its customers.

But there is a hidden cost to this approach. It is much harder to make that data actionable. When you want to apply it to a use case, you end up trying to figure out what data you need and how it works from a huge mass of data.

This is where there is a risk of analysis paralysis; there is simply too much data and too much noise to interpret anything meaningful from the information at hand. Another risk is that there are too many possible conclusions to be derived from the data, and so there is an endless chain of re-evaluating data and discussing alternate extrapolations, until the speed of decision making has been slowed severely.

The costs to the business are very real. Critical use cases fail or end up costing too much to execute.

For example, when behavioral data is used to understand smaller and smaller customer segments or even individual users, the impact of every line of missing data is that much bigger.

Incomplete and inaccurate data leads to the business fundamentally misunderstanding valuable customers. They will be unable to effectively personalize customer journeys, precisely target marketing, or optimize sign-up processes. And they can miss identifying customers who are at risk of churn. All this adds up to costly missed opportunities for the business.
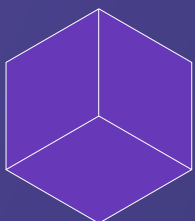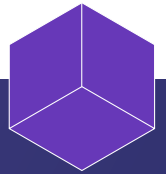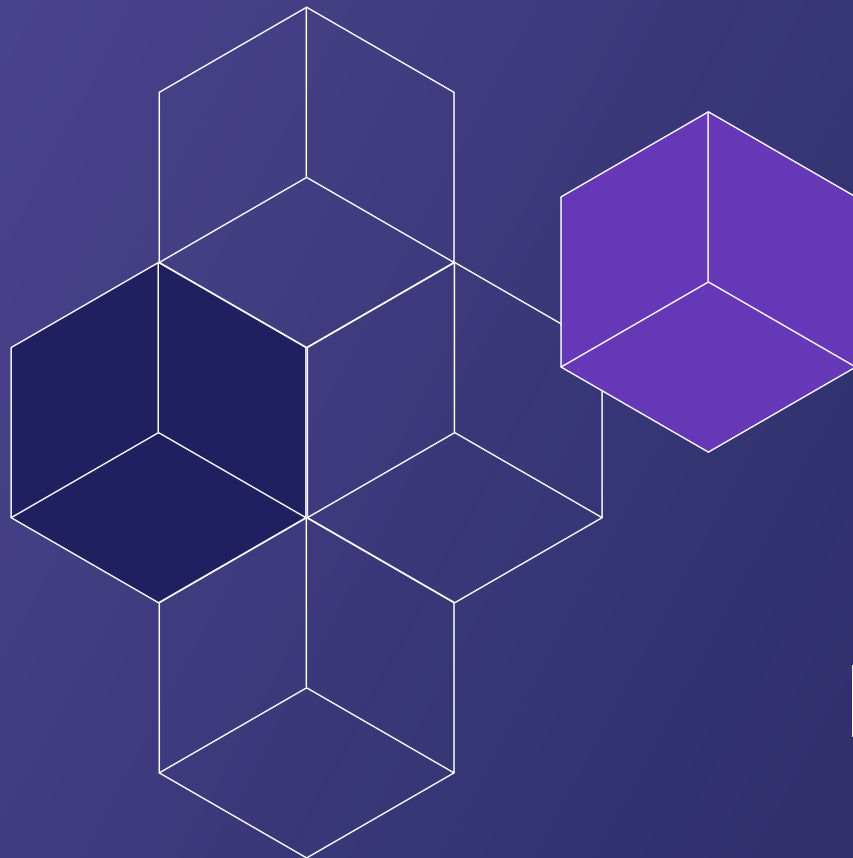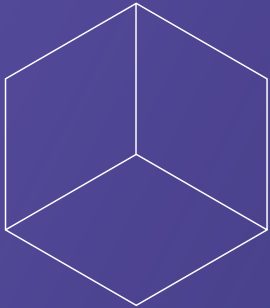
STRUCTURING DATA WITH THE DATA CONSUMER IN MIND

When businesses want to understand their customers and prospects better, they capture event data, which is a record of each action a user or service performs at a given time, with the relevant context.

A page view or logging into a mobile app, for example. When they first start doing this, they don't typically pay much attention to the structure of the data they are collecting.

Some prefer to collect unstructured data because that is the easiest option. It allows them the flexibility to capture whatever data they require and they can easily add more information as their products and services evolve. Others use a predefined data structure that aligns with their packaged analytics solutions—a finite number of columns is specified and can be filled for each event. This makes it simple to use the data in relational databases, initial dashboards, and building visualizations.

The problem with these approaches is that once the business starts to capture bigger volumes of data, data quality can be compromised. Data is collected inconsistently, so that information is hard to find or understand. The burden falls on the data consumer to process it further to make sense of it. And as there is little governance, there is no way for data consumers to ensure that the data they need has been captured. In effect, the business invests in a low quality data asset that its data consumers can't readily action.

REGAINING CONTROL OF HOW DATA IS COLLECTED

To ensure data quality, the data team will need visibility into each stage of their data collection and processing.
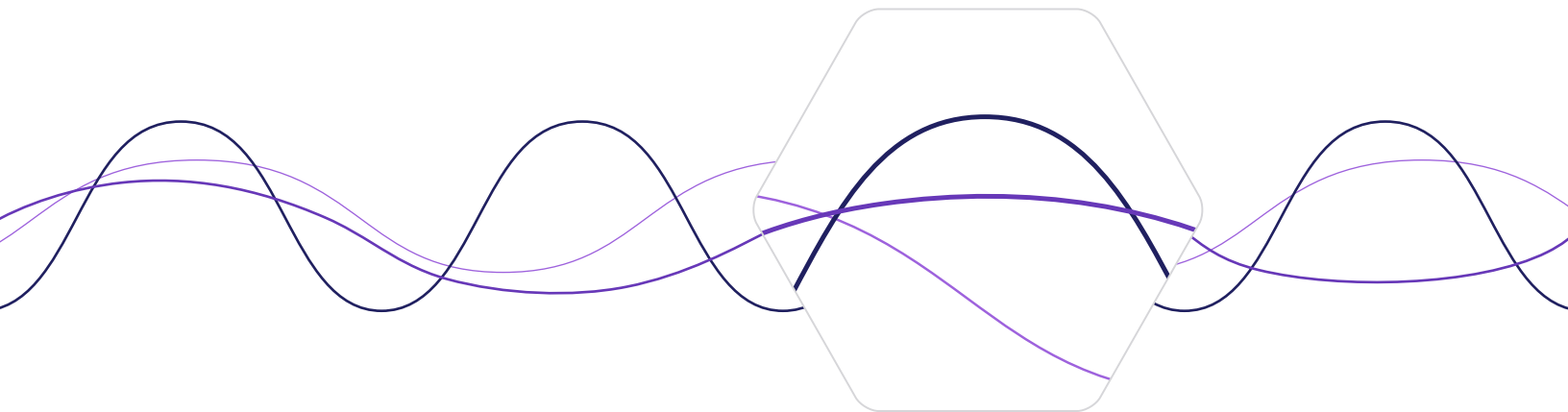
This means having the ability to audit data quality at each stage and validate that no data is missing or inaccurate. Many data collection solutions simply drop bad data or try to transform the data in the data warehouse after the fact. It's near impossible to know what happens to that bad data, and their datasets will remain incomplete. It's far better to surface that bad data so that it can then be correctly handled. This is how data teams can ensure they are collecting trustworthy data sets based on their own needs.

When data teams lose control of their data collection, the wider business will lose faith in that data and in the data team.

And it's very hard to restore that trust. What often happens then is that business leaders revert to gut instinct to make important decisions, or each team independently invests in their own data collection tools. In both scenarios, the business ends up with data silos that obscure its understanding of its customers.

Keeping control of how data teams collect data is also a data governance issue. Consumer trust in organizations is diminishing in response to suspicions about data collection. Reforms that restrict data collection, such as Intelligent Tracking Prevention (ITP) mean that it's now more important than ever that the business controls who can collect or use what data, and how.

It's not just teams within the business who need to trust data—it's also the customers who choose to share their data with the business.

The impact of ITP on data collection is significant.

The aim of the technology is to protect user privacy by preventing third parties with no direct relationship to a website's visitors from tracking them across websites for advertising purposes. But this also affects businesses who use third-party data collection solutions, despite the fact that they use the data to improve customer experience. This data is only usable for seven days, after which ITP removes it. Data becomes unreliable, journey mapping much more difficult, and the business loses insight into its customers over an extended period of time.

DATA COLLECTION RETHINK Q&A

Data collection is as much about common sense and logic as it is about technical capabilities. So a good way to approach it is to start with a set of basic questions that will keep data quality top of mind as the data team establishes its strategy.

Q1   **What are we collecting data for?**

A   Data is increasingly seen as an asset that will help the business deliver its goals. That means data strategy, with data collection at its foundation, will be driven by a business need, such as understanding product performance or usage, or improving the customer experience. Deciding what data the business wants to collect and how starts with a question.

For example, what metrics do we need to measure that will help us understand our customers better? It could be an issue such as identifying where users drop off in your sign-up funnel. Having this focus will help you decide what tools you need to capture these events and ensure the data you collect is accurate and complete across platforms. If you start with your use cases and work backwards from those, it will be much easier to collect the high-quality data you need—and from that, build a data asset that the business can rely on to succeed.

Q2

How will the way you collect your data affect how useful it is?

A

Controlling the way you collect data—by defining and structuring it with your data consumers in mind—allows you to maintain data meaning, data quality, and data governance. These three factors are crucial in deriving real business value from your data.

They ensure that you can socialize the data around the organization as there is a common understanding of what it means, your business trusts the data, and therefore the insights derived from it, and you are able to control what data is collected, and who can access it. How you collect data is fundamental to building a high-quality data asset that creates value for data consumers across the business.

Q3

What tech capabilities will you need to collect quality data?

A

A good start is getting access to your raw data so you can aggregate and model it in a way that best suits your business. Look for a solution that offers you a suite of flexible tools that enable you to collect the data that's most important to you. And remember the impact of restrictions such as ITP and adblockers, where collecting data as a first-party and leveraging server-side tracking means that you can continue to track your users.

Q4 What do your privacy policies mean to your customers, and how does that impact data collection?

A Make sure your data collection ethics are aligned with both regulation and customer expectation. If you can demonstrate transparency and accountability around data collection, customers are more likely to agree to give you their data. The way you secure data, share it, protect its value, and collect it demonstrates your respect for the trust users have shown in sharing their data in the first place. Building your own data infrastructure within your own cloud environment keeps you in control of your data, so that you don't relinquish ownership of something so valuable to a third party that you can't really control. Preserving trust and handling data respectfully is now fundamental to building winning customer experiences.

CUSTOMER INSIGHT

"With Snowplow we are focused on extracting and centralizing data from everywhere, ensuring data quality to be able to stitch everything we need together to get a complete picture. That has required developing a tracking and data mindset in the company from scratch."

Kevin James Parks, Data Engineer, Tourlane