

KEY CONSIDERATIONS WHEN BUILDING PIPELINES FOR BEHAVIORAL DATA

A Guide For Data Engineers

Most companies know that leveraging big data will help them better understand their customers so that they may develop effective marketing and retention strategies. They also recognize the importance of having a single, unified view of each customer's journey, and scaling business use cases with AI and machine learning. That's why a 2019 survey, conducted by strategic advisors NewVantage Partners, suggests that 92% of Fortune 1000 companies reported increasing their pace of investment in big data and AI.

The good news is that almost every company has troves of customer data (both online and offline), to help support their efforts. It's figuring out how to structure this data that's far from simple.

The first step to consolidating customer data is typically to build out a data lake that can tap into all of your internal systems for customer data (i.e., order management systems, CRMs, billing information, etc.). This is where most organizations are today. And while this initiative alone can be complex enough - particularly for companies with tens or even hundreds of different internal systems - it's usually doable with the right investment of time and money.

Next, we have to think about all the behavioral data that's living elsewhere, instrumented through a variety of marketing SaaS vendors such as Adobe, Google Analytics, or email marketing systems. According to research cited by McKinsey, organizations that leverage behavioral data to generate behavioral insights outperform peers by 85% in sales growth and more than 25% in gross margin. And when

we talk about behavioral data, we're talking about marketing SaaS data. There is clearly a ton of value sitting there, just waiting to be tapped.

Generally, the marketing team will be working with many different marketing SaaS vendors to pull data and perform analytics. However, the data is typically not in a usable format out of the gate. Enterprise teams have heavily customized their tools and their customer data (the marketing SaaS vendors have done so as well) and therefore the data formats are all different. The challenge lies in how best to structure the data so that it's useful for the data science team.

"In the past, we've had issues where our clickstream data is in one cloud, and the analyst can't query against other datasets, because it's stored in another database. We were bringing that data in manually, loading it into a database, and then joining on that way."

NICK EDWARDS

Decision Analytics and Insight Manager, Sky

Synthesizing external behavioral data from multiple sources into a uniform and useful format which syncs with existing customer data is an arduous process. Take, for example, a large grocery retailer who decided back in 2016 to build a new analytics platform based on Hadoop, to ingest Adobe clickstream data into their environment. It took them 11 months and they built data pipelines delivering data to Hadoop, but it proved to be unstable, the data quality was not good, and the numbers from Hadoop and Adobe were different and so the metrics were not comparable.

They could run the data feed but not productionize it and, as a result, downstream systems would frequently fail. We've also talked to a Fortune 100 media company who had four engineers spend eight months building a data pipeline from Adobe to their data lake.

Even if your company does have that kind of time and resources, there will likely be other kinds of roadblocks. Every vendor uses different data formats which may change without warning, at any time, and break established processes. Custom code can also break at any time and often these breaks aren't detected until well after the fact, when someone downstream notices something is off. In addition to difficulty monitoring data quality, there are usually high maintenance costs.

For all your behavioral data, what if you could have:

99% match rates with analytics providers

Speed to live

All variables included

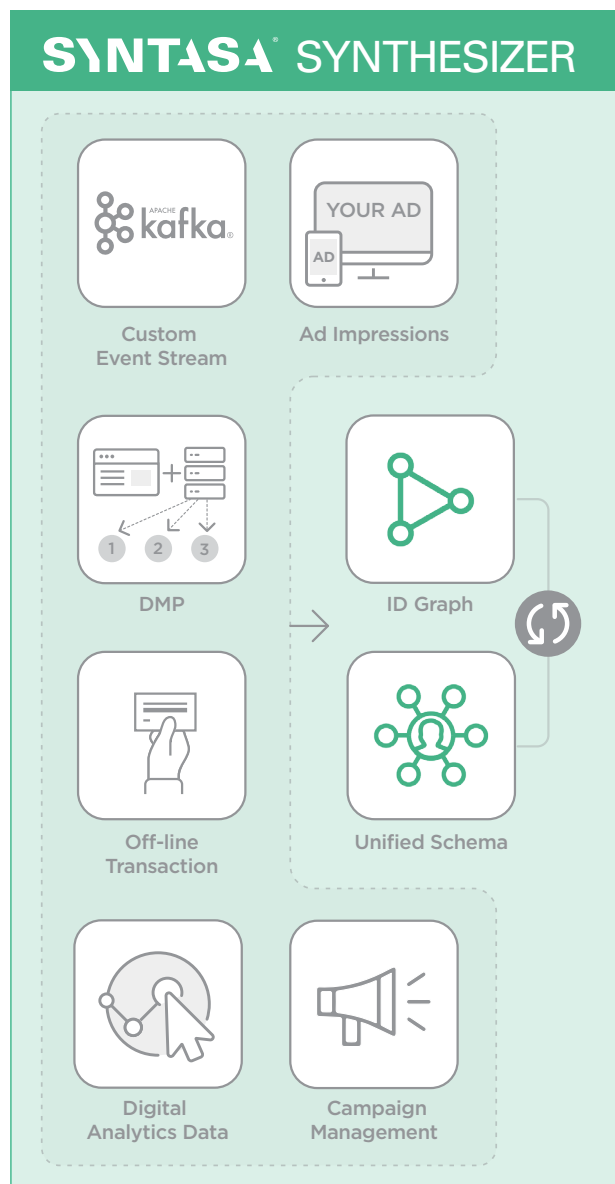
Lookup and product data handlers

Data quality and schema monitoring

Mismatched report suites consolidated

Companies realize they need these requirements, but they still seem hesitant to consider outsourcing the challenge of building something like this. They think it's cost-prohibitive, or they loathe adding yet another external vendor into the mix.

Syntasa was established to make this process as easy as possible for data teams. We've built a product called Synthesizer which is familiar with many marketing SaaS data formats and architects the data in a way that makes it available for immediate analysis and modeling. Keep in mind this is not another ETL tool. Synthesizer is deployed directly within your cloud-based (GCP, AWS, or Azure) or Hadoop environment, and creates scalable production workflows so that experiences can be delivered in real-time.



The speed to market is one of the biggest factors for choosing to work with Syntasa. You'll own your customer and behavioral data, create a custom identity graph, and achieve a unified view of every user's experience. Then you'll be able to focus on those projects you've been wanting to get your hands on, such as building the machine learning and AI-assisted business apps which can take advantage of this newfound behavioral data to increase conversion, retention, and profitability for your organization.



Syntesizer helps build a solid foundation for creating business use case apps, which unlocks more value from your marketing data.