SNOWPLOW**INSIGHTS**

Snowplow Insights **VS** Build-Your-Own Data Pipeline

Do you want your data engineers doing plumbing?
Or using data to differentiate?

Background

In the last few years, it has become increasingly common for companies to invest internal engineering resources into building their data pipelines. This has been driven by:

- Increasing realization of the benefits of owning your own data: especially as companies want to move beyond reporting on it to using it to feed data-driven applications and develop reports that go beyond the standard web and mobile reporting.

- Decreased complexity associated with build-it-yourself: driven by developments in open source frameworks like Apache Kafka and Apache Spark, and web services like Amazon Web Services, Google Cloud Compute, Azure and IBM Bluemixx.

WE BELIEVE THAT SNOWPLOW INSIGHTS OFFERS ALL THE BENEFITS ASSOCIATED WITH BUILDING YOUR OUR OWN EVENT DATA PIPELINE BUT WITH A NUMBER OF SIGNIFICANT ADVANTAGES:

¤ **Deploy your data pipeline fast.** It can take months to build a complete event data stack. We can deploy a complete one for you in a day.

¤ **Use technology that is tried and trusted by thousands of companies.** Snowplow Insights is built on top of our open source technology, that is employed across thousands of companies globally.

¤ **Building good event data pipelines is hard.** Whilst building simple event data pipelines is easy, building good pipelines is hard. There are specific problems that we have solved that you would have to solve if you wanted to build your own pipeline from scratch:

- **How to architect for data quality?** How easy is it to spot (and rectify) when data collection issues arise that impact data quality?

- **How to architect for data reliability?** What happens when your pipeline breaks? How quickly can it be recovered?

- **How do you ensure that your data is easy to consume for end-users?** Event-level data is typically hard for end users like marketers and product managers to work with. How can you package and present that data so that end users can work with it easily and with confidence?

- **How do you evolve your data pipeline with your business as it changes?** How do you ensure your analytics stack keeps up with your business?

¤ **Lean on our technology, team and partner ecosystem.** We provide our customers with round-the-clock support and have a global partner network that can provide additional support to help you get the most out of your data.

SNOWPLOWINSIGHTS

You are right to want to own and control your event data pipeline

Digital event data is a rich asset. Used right, it can be the difference between companies winning and losing. Given that, it is very understandable that a company would want to invest in building its own data pipeline so that it:

¤   Owns all its own data and doesn't have its access to its own data mediated by
    a third party

¤   Can ensure that its data doesn't leave its own environment

¤   Has complete control over what is tracked, from where, and how it is processed
    and used

The trouble is that building, managing, maintaining and evolving an event data pipeline is not easy

Advances in open source big data processing technologies (like Kafka) and cloud technology services (like AWS) mean that building an initial data pipeline has got significantly easier. However building and running an event data pipeline is a significant ongoing commitment. A lot is more difficult than seems initially:

HOW DO YOU MANAGE AND MONITOR DATA QUALITY?

In order to use data to deliver value, people need to trust the data. That means you need:

1. To reconcile the data against any generated by existing analytics solutions / providers
2. A formal framework for understanding the sources of any discrepancies
3. An easy way to proactively monitor for new issues arising in data quality as e.g. new versions of tracked applications are rolled out
4. A path to recovering "bad data"

Snowplow Insights gives you all of these.

HOW DO YOU EFFECTIVELY SOCIALIZE
DATA AROUND YOUR ORGANIZATION?

Event-level data is not easy for end-users e.g. marketers and product managers to consume. Snowplow Insights enables you to define the logic to create a set of derived, modeled data tables that are easy for end-users to consume, and then manages the regular update of those tables with the running of your pipeline over time.

WHAT HAPPENS WHEN YOU WANT
TO CHANGE THE WAY THAT YOU TRACK EVENTS?

No business stands still. Yours will evolve: both as your applications and websites evolve, and also as you become more analytically sophisticated. When that happens, you'll want to update your data pipeline to handle new data, to handle different data and to surface that data differently.

Snowplow Insights is architected from the ground up to enable that evolution: our expectation is that it will always be required. In a DIY pipeline, it is likely to require you to invest data engineering resources to manage every single change request.

WHAT HAPPENS WHEN YOU HAVE AN ENORMOUS TRAFFIC SPIKE?

It should be time to celebrate! But before you do, you need to ensure your data pipeline has scaled to handle the spike.

We have used our experience managing pipelines for more than 100 companies to build a sophisticated pipeline monitoring and autoscaling engine to ensure that traffic spikes are elegantly handled and that real-time pipelines scale down after traffic spikes, so you do not spend on unnecessary resources.

WHAT HAPPENS WHEN SOMETHING GOES WRONG WITH YOUR PIPELINE?

In the real world, things do not work all the time. An outage with a cloud web service provider. A tracking error introduced inadvertently with a new version of your mobile app. These things happen. As a Snowplow Insights customer, when they do:

¤   Our team will be there, either to recover the pipeline (if e.g. it is the result of an AWS outage) or to advise (if it's a data collection issue that needs to be addressed)

¤   We have extensive experience dealing with all manner of failures, accumulated over five years

¤   Our Support and Tech Ops teams are globally distributed and available around the clock

Snowplow Insights offers a wealth of other benefits...

QUICK TO DEPLOY

Deploy in a day vs spending months/years building, testing and evolving you own event data stack.

OUT-OF-THE-BOX SUPPORT FOR TRACKING DATA FROM AN ENORMOUS NUMBER OF PLATFORMS AND CHANNELS

We have an enormous library of SDKs and third party integrations so you can track a host of data beyond web and mobile, including email, push notification, display ads, support system, marketing attribution, call center, smart car, smart home, smart TV... The list goes on.

ENRICH YOUR DATA
WITH 1ST AND 3RD PARTY DATA

It is not always possible to track all the relevant data points at the point of data collection. Fortunately, Snowplow Insights has a plug-in enrichment module, enabling you to configure common enrichments like:

¤   Weather

¤   Geolocation

¤   Marketing attribution

¤   Referer parsing

¤   Currency conversion

In addition, you can enrich your event stream with data held in any SQL database or via any RESTful API - for example - your own customer or product metadata sets, ensuring your data is as rich and granular as possible to power the broadest range of analytics application.

TRIED AND TRUSTED

Snowplow Insights is trusted by close to 100 different companies. Our open source technology, on which it is built, is trusted by 1000s of different companies around the world.

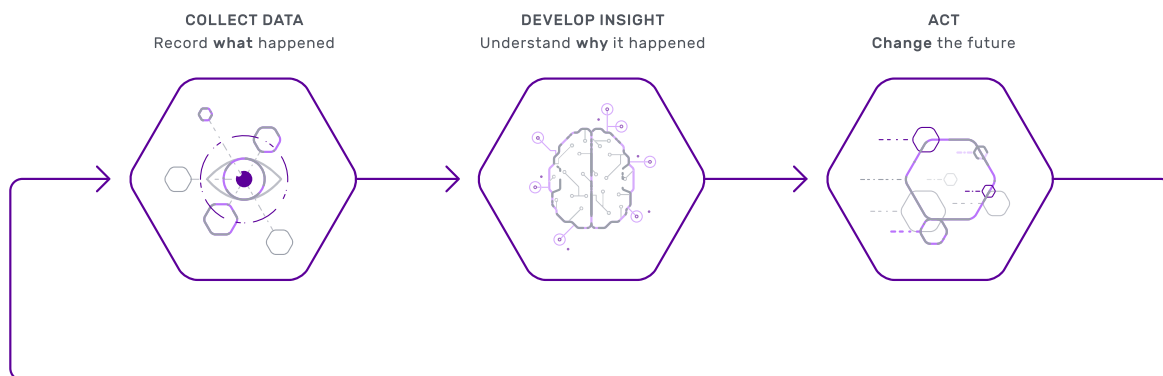## A RAPIDLY DEVELOPING PROPOSITION

When you buy into Snowplow Insights, you don't just buy into its feature-set today. You buy into a product with a long and growing roadmap of developments including support for:

¤ New (non-AWS) cloud providers, including both Google Cloud Compute and Microsoft Azure
¤ Streaming data into more storage targets
¤ Modeling data in-stream, in real-time

A GROWING ECOSYSTEM
OF EXPERT KNOWLEDGE, HELP AND ADVICE

Work with our Data Analytics and Data Engineering teams and with our growing global partner network, to get more from our technology and your data.

...and all the benefits of build-your-own

COLLECT DATA
Record **what** happened

DEVELOP INSIGHT
Understand **why** it happened

ACT
Change the future

¤   Your data, under your control, in your own cloud environment

¤   Complete freedom to decide what you want to track, from where, and how you want to structure that data

¤   Complete freedom to decide how you want to enrich, process and surface that data in different dashboards, reports, predictive models and other data-driven applications

SNOWPLOW INSIGHTS

DATA COLLECTION DONE RIGHT

Snowplow Insights puts you in control of your data collection. You decide what you want track. We deliver that data to you, in your own data warehouse, so you can ask any question of your data, perform any analysis, use any analytics tool. We surface your data to you, in real-time, so you can act on it.

Set your team free to use data to differentiate.

SNOWPLOW

DIFFERENTIATE WITH DATA