SNOWPLOW INSIGHTS

COMPARISON WHITEPAPER

Snowplow Insights **VS** SaaS load-your-data warehouse providers

We do data collection right.

Background

We were the first company to launch a platform that enabled companies to track event-level, web and mobile data directly into Amazon Redshift.

Today the value of warehousing your digital event data in your own data warehouse is well understood and a large number of suppliers have built solutions to enable this, including:

- **Alooma** (www.alooma.com)
- **Astronomer** (www.astronomer.io)
- **ironSource Atom** (www.atomdata.io)
- **Tealium** (www.tealium.com)
- **TreasureData** (www.treasuredata.com)
- **Segment** (www.segment.com)

There are some very significant differences between Snowplow Insights and all the above solutions. These matter, because data collection is the foundation of the data value chain. We'll explain those differences in the next pages.

Why Snowplow

Data quality

To be transformative with data, people need to trust that data. Snowplow is architected from the ground up to enable users to build high confidence in the data:

- **Auditable.** Each stage in the Snowplow Insights pipeline is fully auditable. You have direct access to the raw collector logs, the output of the validation, enrichment and data loading processes, to validate that every event hitting the pipeline is processed.

- **A formal approach to handling bad data.** Bad data is a fact of life. Snowplow Insights has a formal validation step: where input data is validated against a schema to ensure it is correctly formatted and will be processed successfully. In the event that it does not, we surface the data, with the associated error messages, to a dedicated location so our users can proactively monitor and easily spot if data quality issues emerge. Where they emerge, users can address the issue upstream and have the opportunity to reprocess the 'bad' data. In addition, we have a growing toolkit for spotting implementation issues that result in bad data before these are pushed to production.

- **Non-lossy pipeline.** All our competitors (both the Saas load-your-datawarehouse vendors and Enterprise Digital Analytics vendors) have lossy pipelines in the sense that if you send data that is not successfully processed, it is simply dropped. In contrast, the Snowplow Insights pipeline is non-lossy: if there is an issue processing the data, we surface it (as discussed above) rather than bury it.

Why Snowplow

Data richness

The richer an event data point is (in terms of the number of associated data points), the more questions that can be answered with that data point.

Snowplow Insights is built to deliver rich data:

- **Automatic tracking of rich contextual information with events recorded from different platforms.** For example any event that you record from the web can be automatically recorded with data about:

    - Who performed the action (based on a 1st and 3rd party cookie ID, IP address, browser fingerprint)
    - Additional user or device identifiers from third parties e.g. from Google Analytics, Augur
    - The web page (page URL and title)
    - The referer that drove the user to that page
    - The speed with which the web page loaded
    - A/B testing data from third party providers (e.g. Optimizely)

DATA RICHNESS (CONT.)

- **Data is enriched before being loaded into the data warehouse.** Often there are limits to what data is available at the place where the event is recorded. Snowplow Insights includes a formal enrichment module, where event data can be enriched with additional first and third party data before being loaded into the data warehouse. That means, for example, if you record that a user with ID 123 viewed video ABC, you could pull in additional metadata about both the user 123 and the video ABC. Enrichments include:
  - SQL Enrichment: pull in additional data from your own databases
  - API Enrichment: pull in additional data from your own and third party APIs e.g. Clearbit
  - IP Enrichments: pull in geolocation, business and connection information data from different Maxmind databases
  - Campaign Attribution Enrichment: identify which paid campaigns drove users to your website or mobile app, and append the appropriate campaign metadata to the event
  - Current Conversion Enrichment: calculate monetary values in different currencies
  - Weather enrichment: identify what the weather was like when the event occurred
  - Cookie Extractor Enrichment: automatically fetch additional cookie data

Why Snowplow

Making the data easy to query and consume

We have put a lot of work into ensuring that the data is easy to consume, making it more likely that our users will get value out of the data.

- **Easy to understand data structure.** Data is loaded into tidy tables in the data warehouse.

- **Predictable data structure.** As you evolve your tracking (as your digital proposition evolves and the questions you ask of your data evolves) the structure of the data in the data warehouse evolves in a completely predictable way, so you always know where to find individual data points.

- **A formal approach to data modeling.** Event-level data is not easy to work with. In order to socialize that data successfully around an organisation, it needs to be modeled. Snowplow Insights includes a well developed framework and set of templates for building out that essential modeling process on the data once it has been loaded into the data warehouse.

Why Snowplow

Evolve your analytics stack with your business

Businesses change. So your analytics stack will need to change with you, specifically as:

1. Your digital propositions (e.g. mobile apps and website) change. As they do, the events that you'll record from them, and the associated data points, will need to change.

2. You become more analytically sophisticated. As you do, you'll want to capture increasingly granular data.

Snowplow Insights is built from the ground up to elegantly manage this evolution process. We have a formal process for:

- **Updating the structure of the data tracked.** This means that structure can change over time, without users ever having to either lose old data, or have data turn up in surprising tables and columns in their data warehouse.

- **Updating the data model.** This ensures that companies can easily manage updates to the structure of the data in the tables that most end-users interact with.

Why Snowplow

Real-time

As well as regularly loading data into your data warehouse (up to every hour), Snowplow Insights can deliver your validated, enriched data to you in near-real-time directly into your own Amazon Kinesis stream and Elasticsearch cluster. This means you can build real-time dashboards and data-driven applications (e.g. personalization, recommendation, marketing automation) that consume and act on the data in real-time.

Why Snowplow

Total control with your own pipeline

With a SaaS provider, you share a data pipeline with every other customer of that SaaS vendor.

With Snowplow, you have your own complete data pipeline, in your own AWS account. That means you have total control over your own data, which never leaves your own AWS account. (Unless you remove it yourself.)

It means you can decide how frequently you want to run the pipeline, how much latency you want to tolerate, how much you want to spend on your infrastructure each month. All whilst enjoying all the benefits of a serviced offering: we take care of delivering your data to you, so you can focus on using the data to make a difference.

Why Snowplow

Cost

Snowplow Insights is charged on a fixed monthly fee basis. We do not penalise companies for tracking enormous volumes of data. We enable and encourage them to!

As a result, we are typically an order of magnitude cheaper than our competitors.

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

snowplowanalytics.com