Data Flywheels

- What are they?
- Why you should be building them?
- How you should be building them?

Date

© 2024 SnowplowData Holdings, Inc. All Rights Reserved.

My name is Yali, I am co-founder and CTO at **Snowplow**



About me

- Data guy
- Co-founded Snowplow as an open source project in 2012

About Snowplow

- Customer Data Infrastructure: create rich, proprietary, first-party customer data directly in your data platforms
- Killer apps: personalization, recommendation
- 1000s of open source users, 250 commercial customers globally

© 2024 SnowplowData Holdings, Inc. All Rights Reserved.

Before I start, can we do a quick show of hands?

Data warehouse or data lake?







© 2024 SnowplowData Holdings, Inc. All Rights Reserved.

Before I start, can we do a quick show of hands?

Data warehouse or data lake?







Real-time streaming?





Cloud Pub/Sub

© 2024 SnowplowData Holdings, Inc. All Rights Reserved.

Before I start, can we do a quick show of hands?

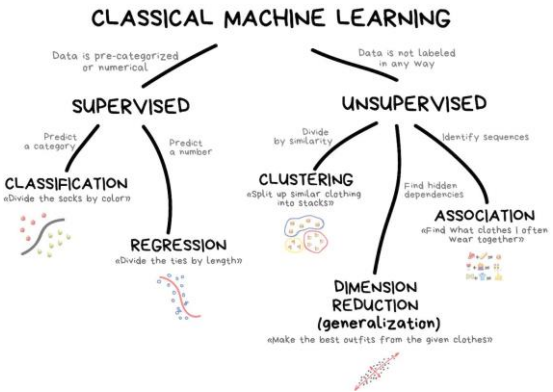
Data warehouse or data lake?

snowflake

databricks

BigQuery

Classical ML in production?



Real-time streaming?

kafka

Cloud Pub/Sub

© 2024 SnowplowData Holdings, Inc. All Rights Reserved.

Before I start, can we do a quick show of hands?

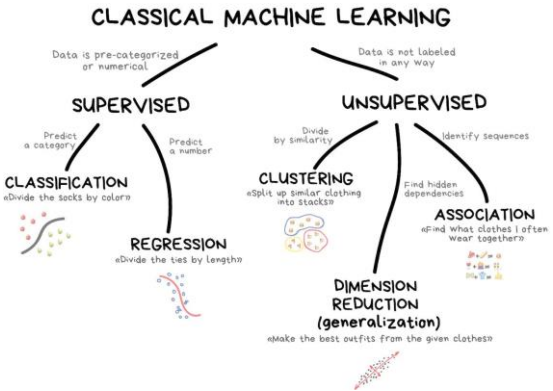
Data warehouse or data lake?

snowflake

databricks

BigQuery

Classical ML in production?



GenAI pilots?



Real-time streaming?

kafka

Cloud Pub/Sub

© 2024 SnowplowData Holdings, Inc. All Rights Reserved.

Before I start, can we do a quick show of hands?

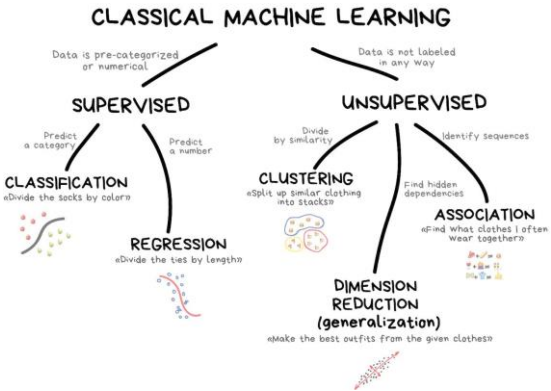
Data warehouse or data lake?

snowflake

databricks

BigQuery

Classical ML in production?



GenAI pilots?



Real-time streaming?

kafka

Cloud Pub/Sub

GenAI in production?

© 2024 SnowplowData Holdings, Inc. All Rights Reserved.

I was lucky to hear Jensen Huang talk about the “Data Flywheel” at both Databricks and Snowflake Summits



with Sridhar Ramaswarmy at Snowflake Summit



with Ali Ghodsi at Databricks Summit

© 2024 SnowplowData Holdings, Inc. All Rights Reserved.

Jensen Huang can’t stop talking about data flywheels

“The important thing about the chat bot, the customer service, is partly about the fact that you could automate. But it's ***mostly about the data flywheel****. You want to* ***capture the conversation****.* ***You want to capture the engagement in your data flywheel****. It's going to create more data, of course. We're probably seeing data expanding about growing about 10X every five years. Wow.”*



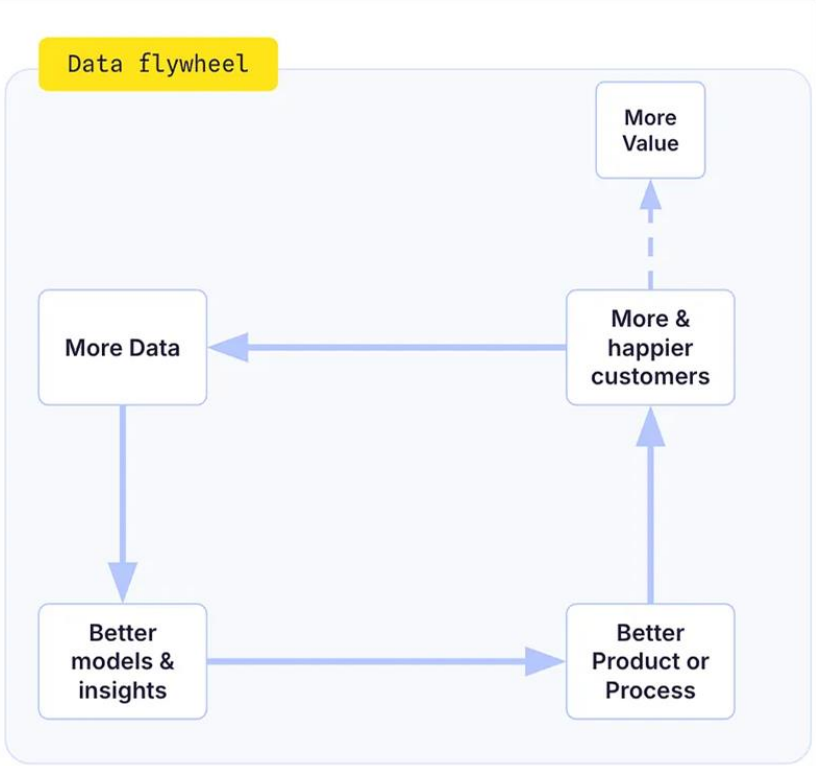
Jensen Huang
CEO, NVIDIA

© 2024 SnowplowData Holdings, Inc. All Rights Reserved.

But what exactly is a data flywheel?

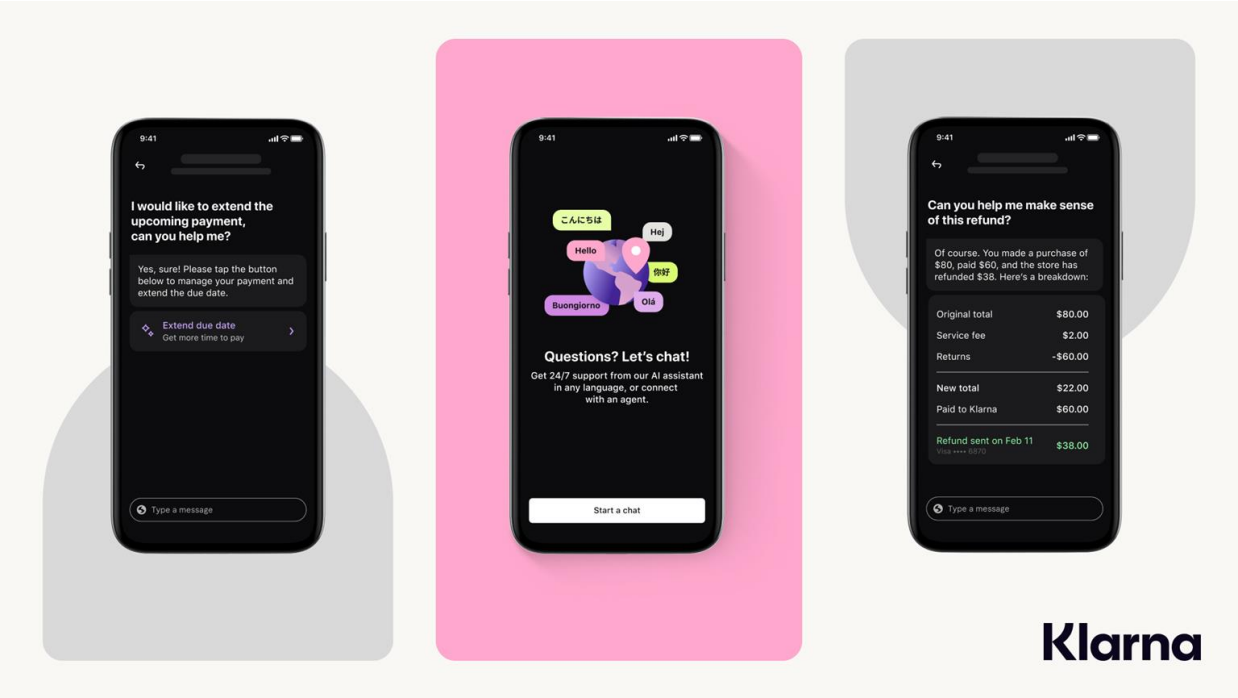
- **Data flywheel**: a positive feedback loop that drives sustained and growing value
- With customer data, there's a flywheel where more data drives improved products and services, which attracts more customers, which creates more data

amazon



© 2024 SnowplowData Holdings, Inc. All Rights Reserved.

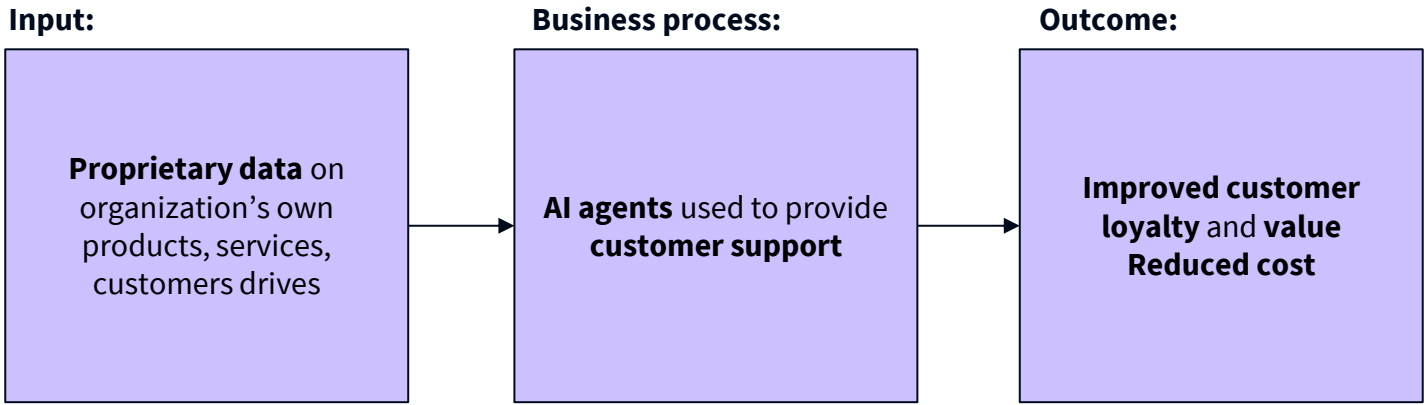
But Jensen’s talking about a different dynamic. Let’s look at the Klarna support agent example



- Replaced 700 agents
- Speaks more than 35 languages
- Resolves issues faster (2 mins vs 11 mins)
- On par with human agents (satisfaction score)
- $40M profit saved

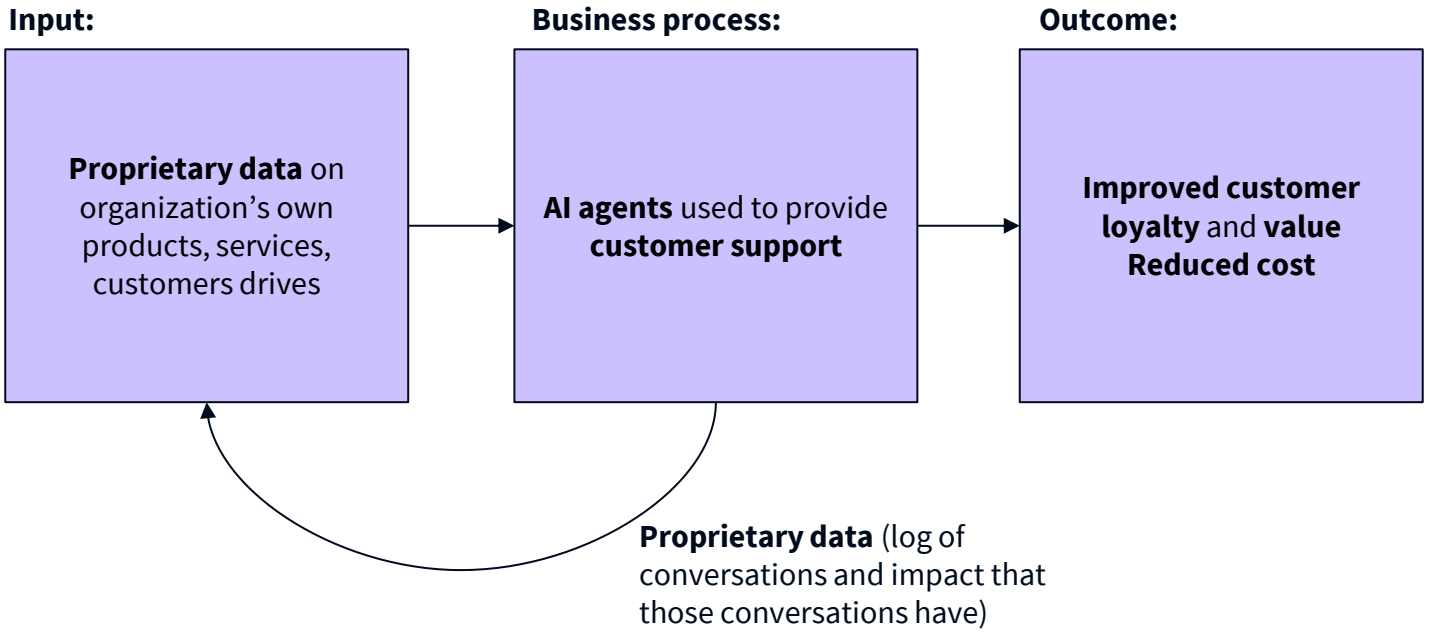
© 2024 SnowplowData Holdings, Inc. All Rights Reserved.

How is the support agent driven by a data flywheel?



© 2024 SnowplowData Holdings, Inc. All Rights Reserved.

The flywheel outputs proprietary data, that serves as input to drive improved results



© 2024 SnowplowData Holdings, Inc. All Rights Reserved.

These flywheels are everywhere



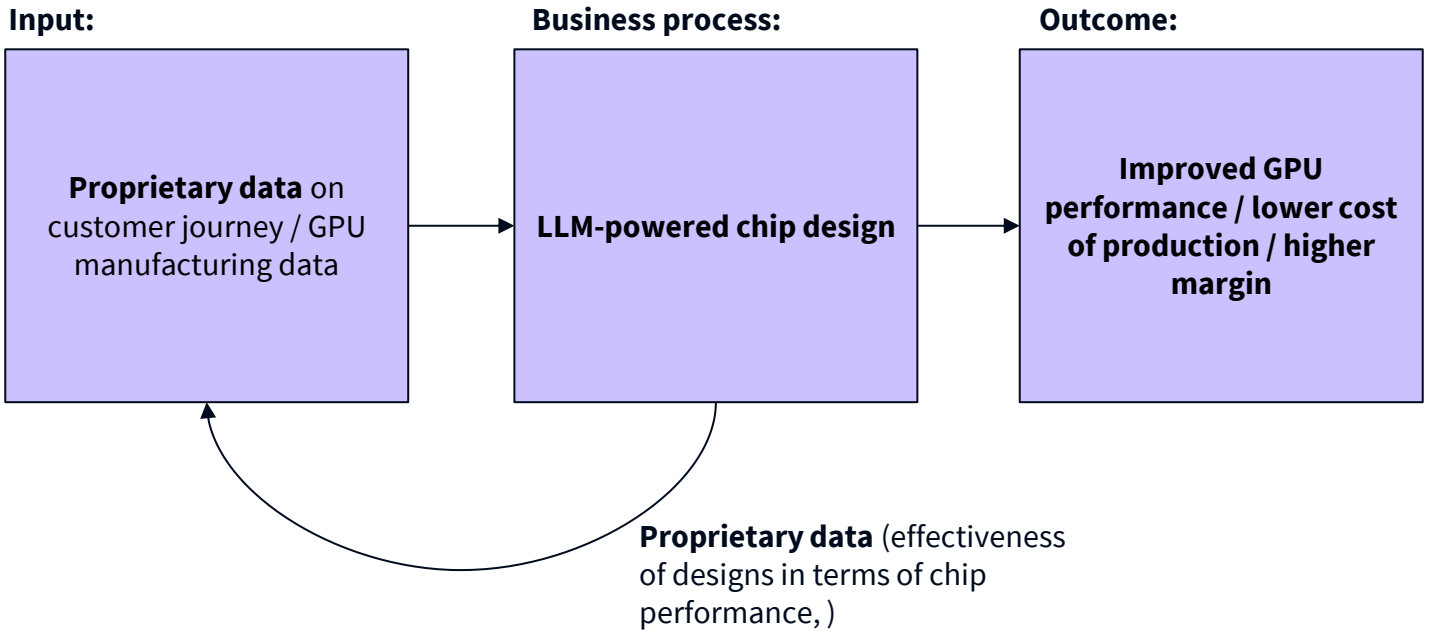
*All of us have **flywheels of data** that we've been collecting for a long time. We're sitting on mountains of it. …*

*Yeah. And we're using it in our chip design. We're using it in our bugs database. We're using it in creating new products and services. And we're using it in our supply chain. And for the very first time, we now have **an engineering process that starts with data processing and refinement and learning models and then deploying the models and connecting that flywheel**, collecting more data. Isn't that right? Yeah.*

”

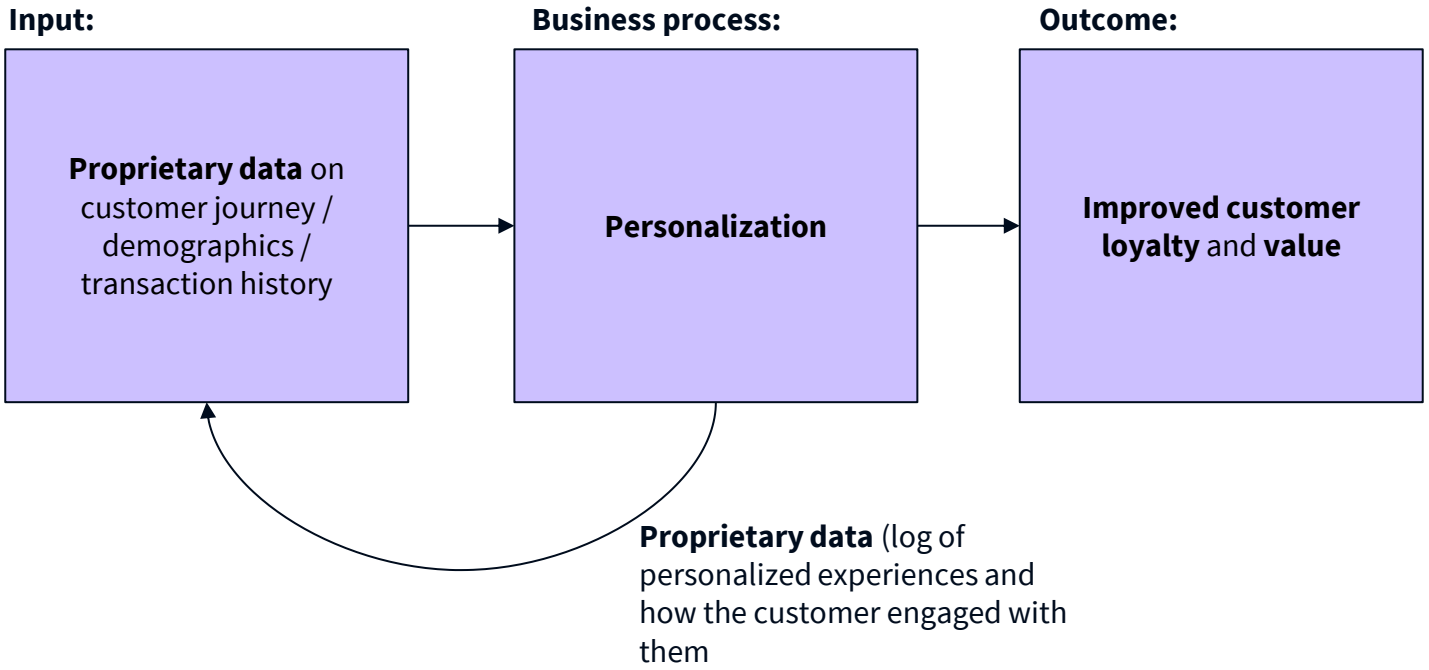
© 2024 SnowplowData Holdings, Inc. All Rights Reserved.

These flywheels can be created anywhere there is an agentic business process



© 2024 SnowplowData Holdings, Inc. All Rights Reserved.

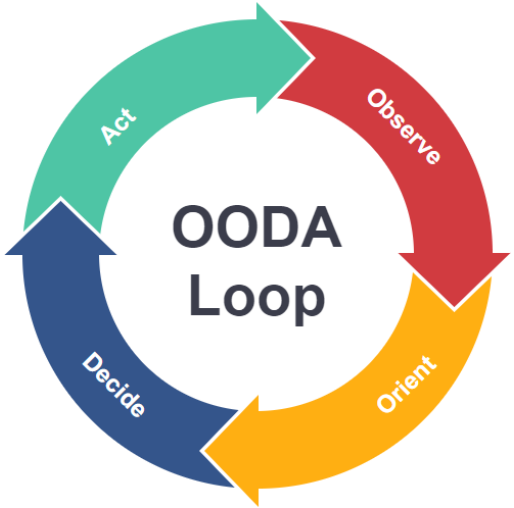
And anywhere there is an ML powered process



© 2024 SnowplowData Holdings, Inc. All Rights Reserved.

These flywheels drive *sustained* competitive advantage. They are examples of OODA loops





F-86 Sabre. John Boyd wanted to understand how US pilots were winning more dog fights than their Russian adverseries in their Mig-15s, even though the Mig-15s were technically superior. The answer? The bubble canopy + hydraulic controls i.e. better observation and faster action)

© 2024 SnowplowData Holdings, Inc. All Rights Reserved.

What's most important in an OODA loop - and in a data flywheel?

1.

Complete the cycle

If the cycle breaks at any point, you lose the ability to **feed your learnings back** into the next go-round

2.

Speed of the cycle

As John Boyd learnt in Korea, speed through the loop is how you outcompete. You are learning and **adapting** to your environment faster

3

Scale of the cycle

The more data you collect, the more insights you drive into your business. You benefit from the **compounding** of knowledge you can apply to your business

But ***how*** do you go about building these data flywheels?



© 2024 SnowplowData Holdings, Inc. All Rights Reserved.

This is a question we've spent a lot of time on at Snowplow



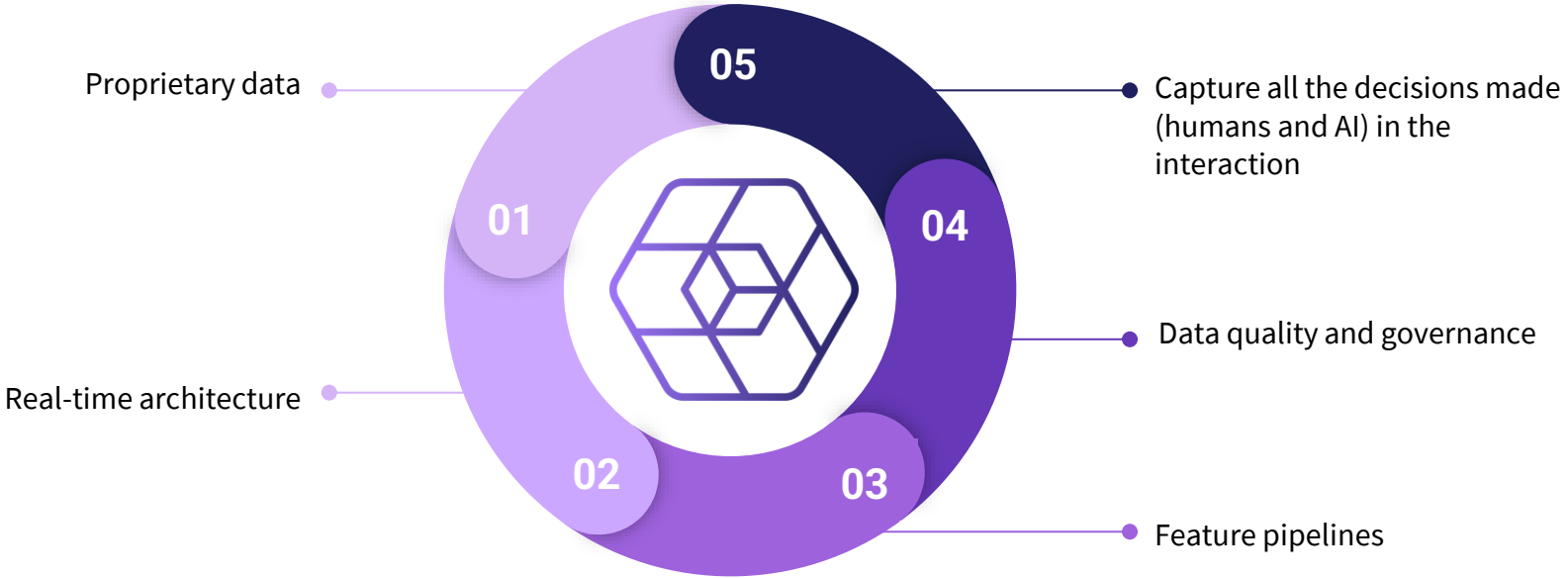
- Attribution
- Conversion rate optimization
- Customer segmentation
- …

- Recommendation
- Personalization
- Fraud and bot detection
- …

- Data Agent
- Customer agents
- Marketing agents
- …

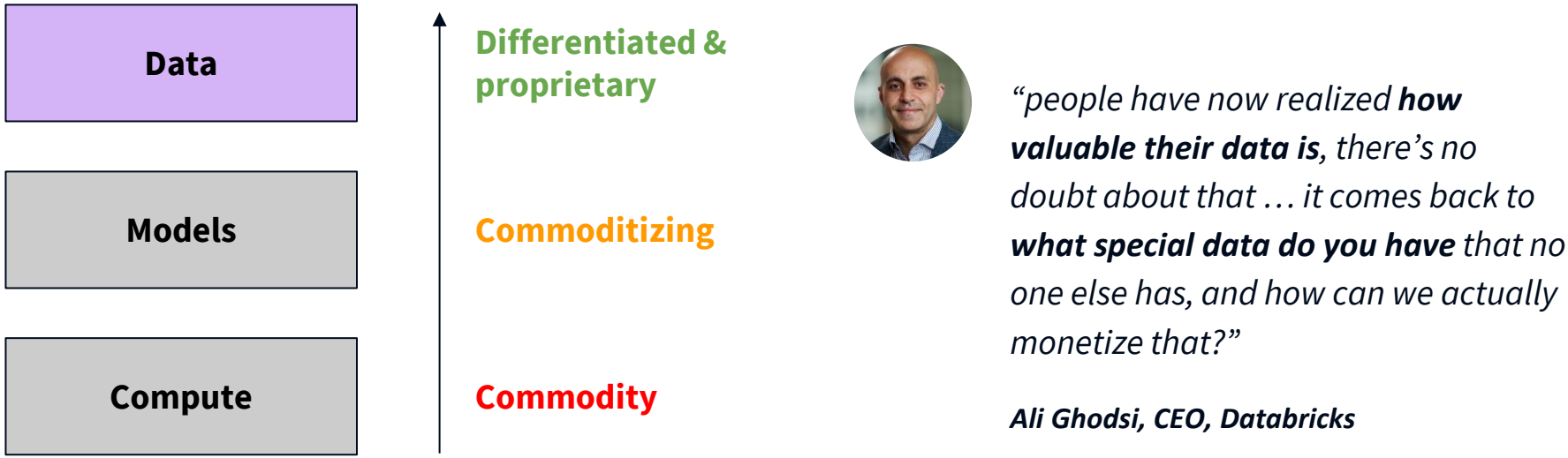
© 2024 SnowplowData Holdings, Inc. All Rights Reserved.

Five key elements to **building data flywheels**



© 2024 SnowplowData Holdings, Inc. All Rights Reserved.

Proprietary data: this fuels the data fly wheel



*"people have now realized **how valuable their data is**, there's no doubt about that … it comes back to **what special data do you have** that no one else has, and how can we actually monetize that?"*

Ali Ghodsi, CEO, Databricks

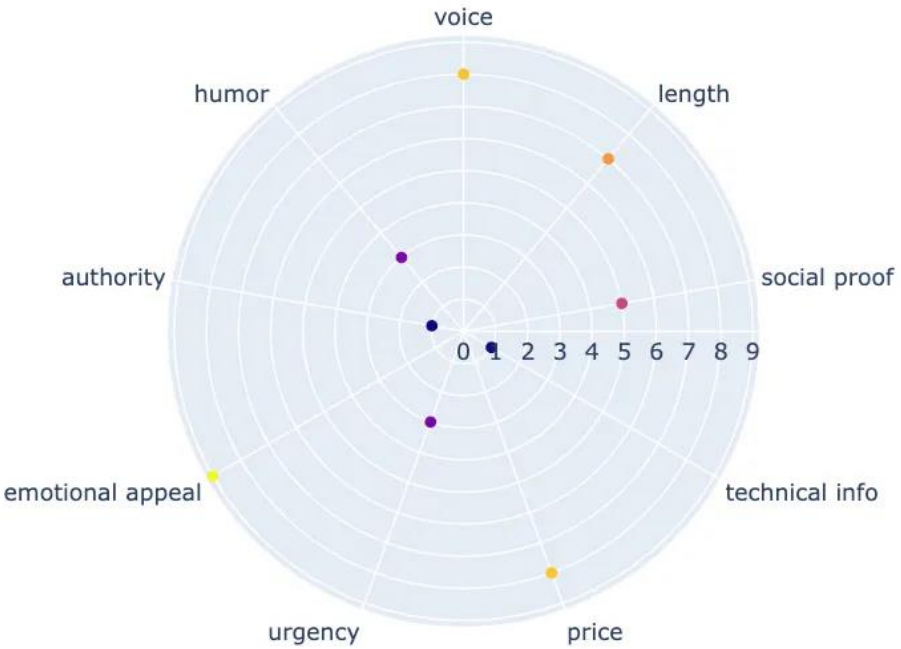
© 2024 SnowplowData Holdings, Inc. All Rights Reserved.

Real-time architectures

Consider not only the time to serve but also the time to learn



Feature pipelines



An example polar plot of a users preference across 9 different dimensions.

Content that emphasis active voice and emotional appear whilst avoiding technical details is likely to resonate best with the user.

© 2024 SnowplowData Holdings, Inc. All Rights Reserved.

Data quality and governance

Quality and governance need to be built into the data manufacture process, not bolted on as an afterthought



I know what this data *means*



I *trust* that it is right

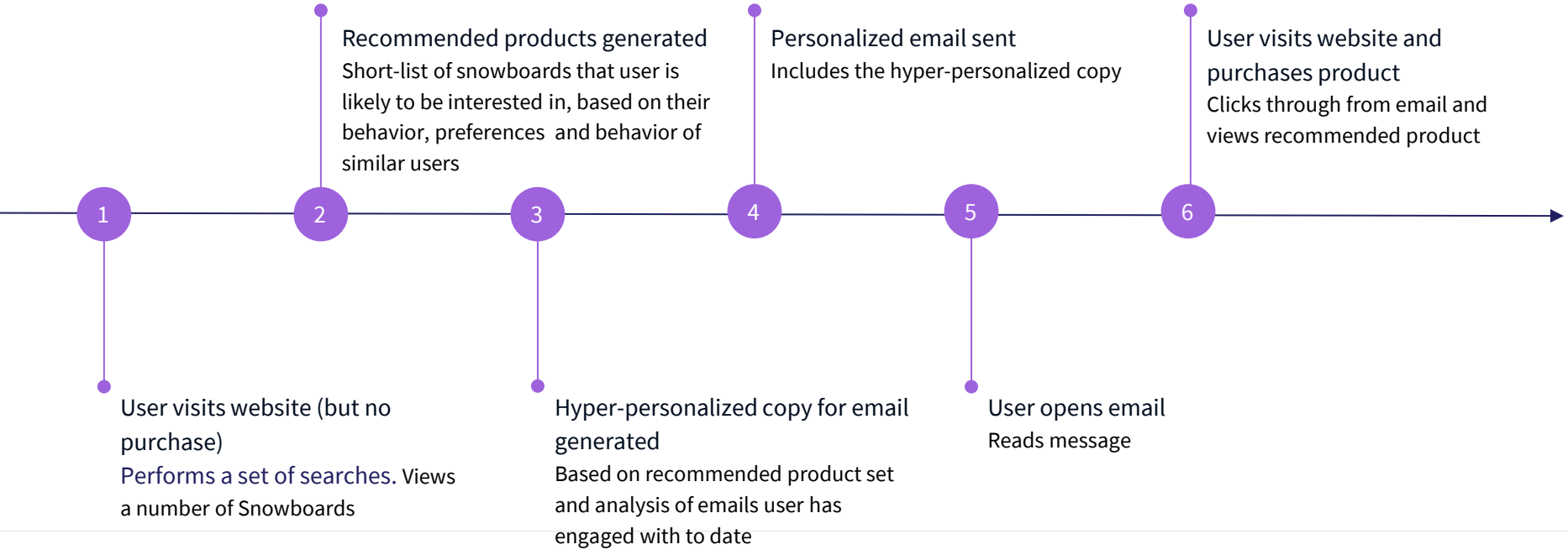


I know that this use of the data is compliant

© 2024 SnowplowData Holdings, Inc. All Rights Reserved.

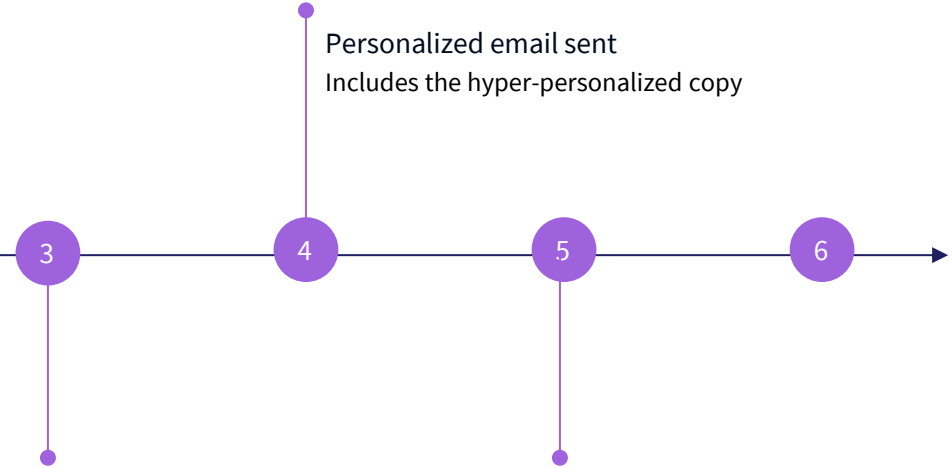
Capture all the actions taken and decisions made (for the humans *and the* AIs)

Example application to re-engage shoppers with hyper-personalized emails



© 2024 SnowplowData Holdings, Inc. All Rights Reserved.

Example payload for AI-driven event

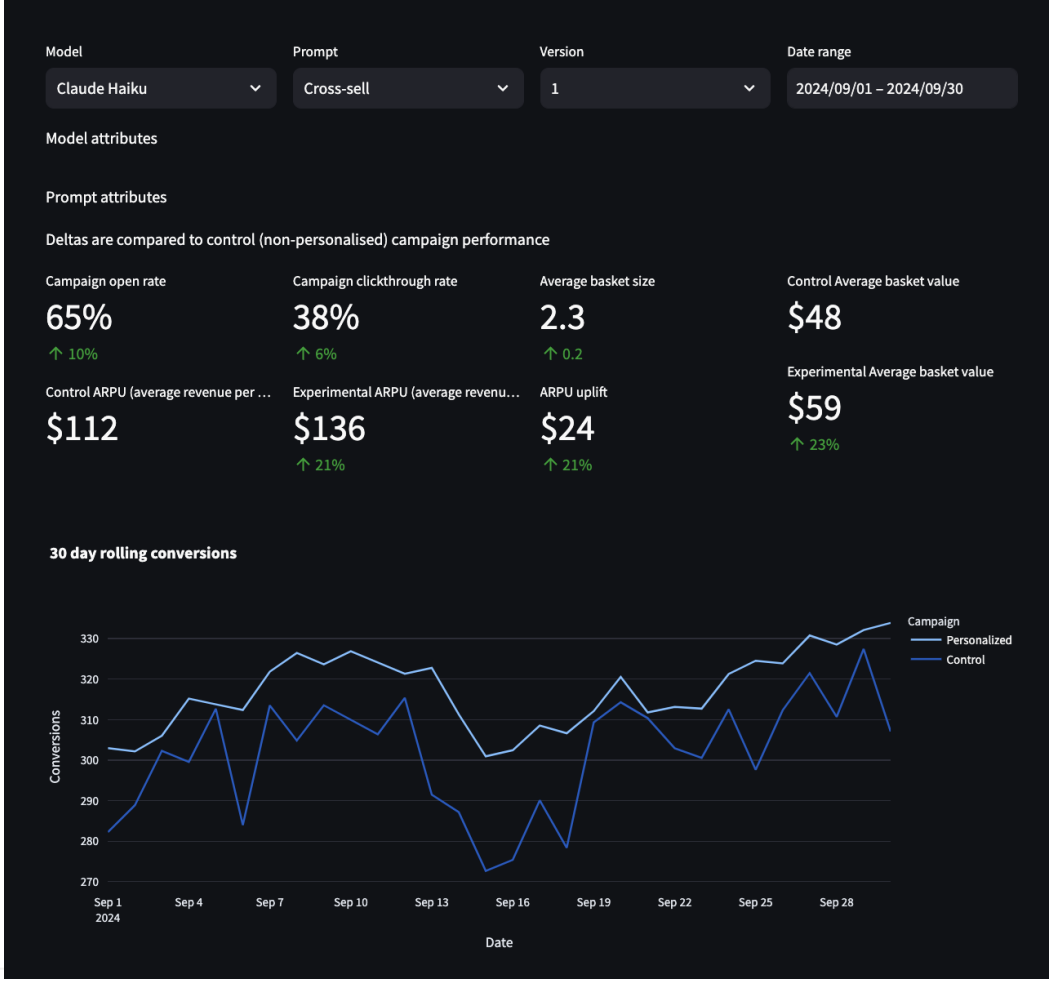



```
{
    "event": "personalized_mail_copy_generated",
    "campaign": {
        "id": "f2180d12-...",
        "copy_generated": "Unleash Your Powder Passion
        with the Ride Wild Life Snowboard! Are you tired
        of battling through the deep snow, longing for a
        board that effortlessly carves through the
        powder? Look no further than the Ride Wild Life
        Snowboard!..."
    },
    "customer": {
        "email": "jane@snowstore.com",
        "profile": {
            "snapshot_time": "2024-10-09T14:10:256",
            "user_features": {
                "active_voice": 5,
                "price_sensitivity": 8,
                ...
            },
            "marketing_consent": true,
            "basis_of_processing": ["consent"]
        }
    }
    "model": {
        "service": "bedrock",
        "provider": "anthropic",
        "model_id": "claude-3-haiku-20240307-v1.0",
        "version": "2",
        "system_prompt": "You are a ...",
        "user_prompt": "Generate a ...",
```

SNOWPLOW

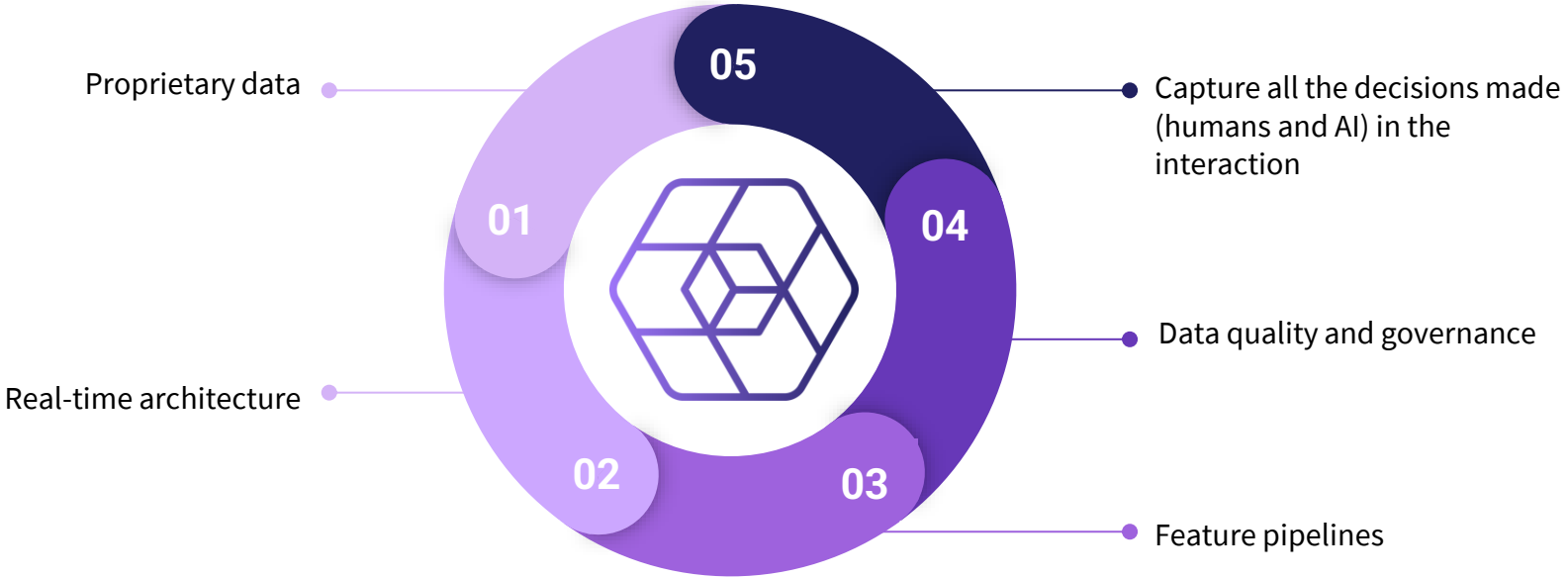
© 2024 SnowplowData Holdings, Inc. A

This data enables us to measure and learn



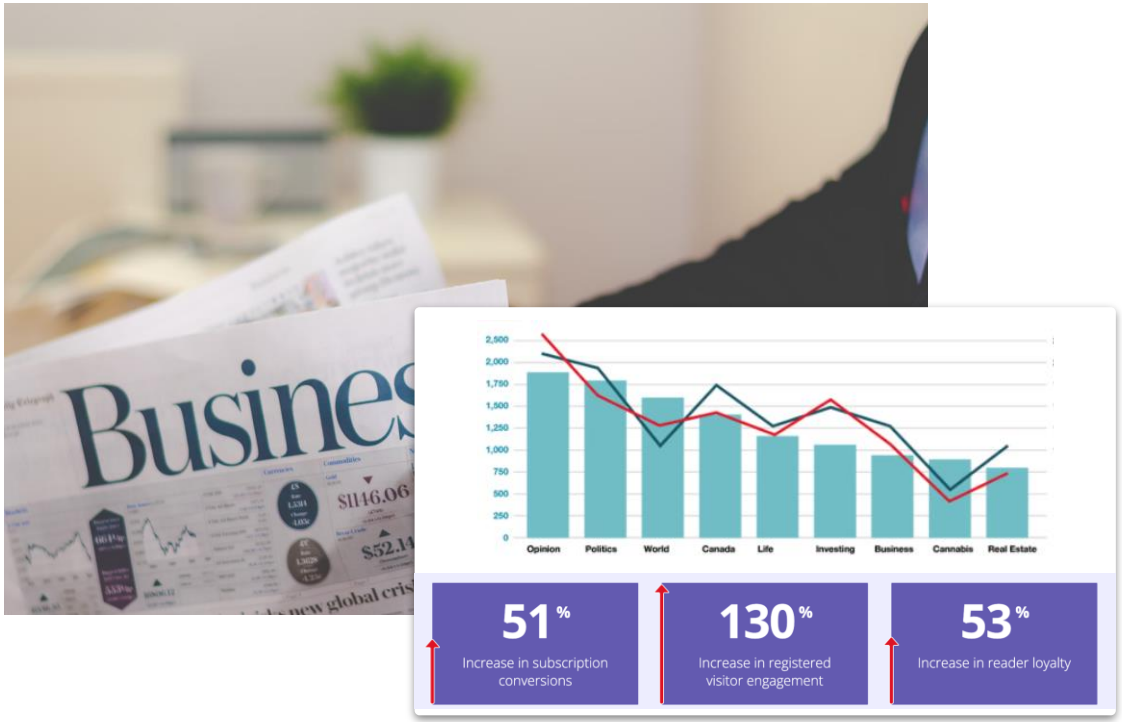
© 2024 SnowplowData Holdings, Inc. All Rights Reserved.

Building these flywheels is **very doable**



© 2024 SnowplowData Holdings, Inc. All Rights Reserved.

And incredibly impactful



- $ driven per content item
- $ impact of placement
- Automate placement of content to maximize $
- …
- Personalized paywalls

© 2024 SnowplowData Holdings, Inc. All Rights Reserved.

Learn more about how we've partnered with different companies to build data flywheels

Visit us at our **booth**

© 2024 SnowplowData Holdings, Inc. All Rights Reserved.