Server Side Tracking with Snowplow

Paul Ferrett
Head of Development, Catch Group
Snowplow Analytics Sydney Meetup - June 25, 2019

A brief history of data at Catch

- Not so long ago we used to do things like most others...
  - **Google Analytics** for website tracking
  - **Custom reports** in our back office systems
  - **TM1** for sales reporting, planning, and forecasting
  - **Bespoke** prediction models for stock movement and warehousing
  - Daily **email reports** from the “Reporting Server(s)”
  - Rely on dashboards within **individual business systems** (Campaigns, marketing, personalisation)
- Everything grew ad-hoc, and there was no centralised “BI” team or platform
- *… and rarely did they communicate the same numbers!*

We do things better now

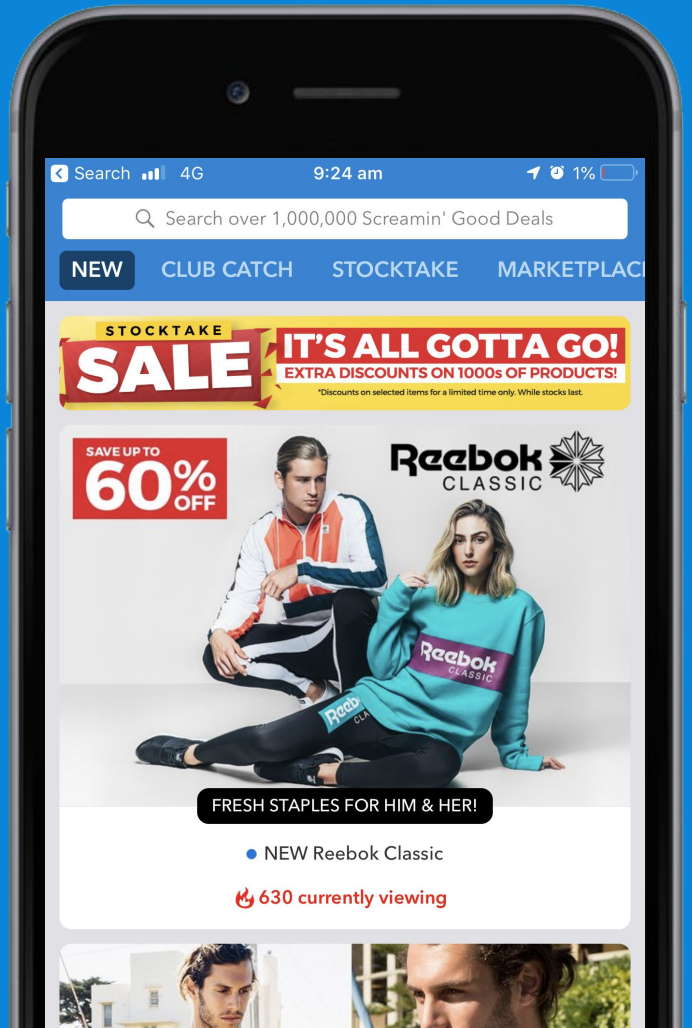
- Centralised BI team
- Data lake in S3/Redshift
  - Snowplow Batch Pipeline
  - External data loaded in
  - Internal databases feed in
- Dashboards in Periscope
- *...and people almost never look at GA!*

Quick Snowplow stats

- Last 12 months
  - 3TB data collected
  - 25B rows
- Daily growth
  - 500k sessions
  - 95M records
- Monthly growth
  - 2.9B rows
  - 350 GB

Why Server Side Tracking?

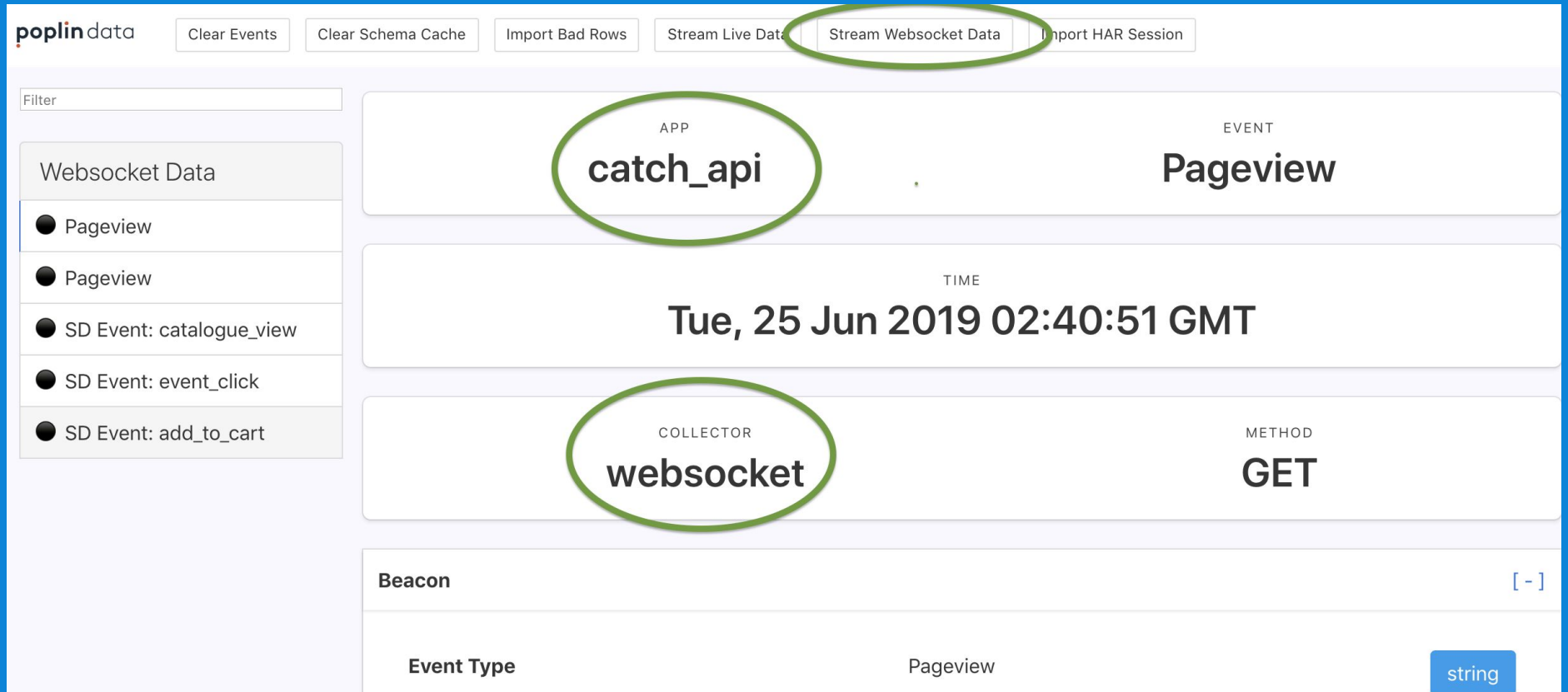
- Avoid sending context data via API
- Decouple tracking changes to app releases
  - Started tracking the app with no release
  - force-update = uninstalls
  - Able to make schema changes more often
  - Fix data issues on the API side
- Multiple platforms with less work ...one day



Our Server Side Tracker

- Built from scratch in PHP
- Emits data to the collector after client request complete
- Uses the same context building code as the front end
- `/tracking` endpoint for interactions
- Pluggable Emitter
  - Curl; Send to the Snowplow Collector
  - Logfile; Simple debugging
  - Event bus; Websocket connector, retryable queue for collector errors, Microservice consumers

Our biggest problem; Verifying server side data sent to the collector!



Benefits of **Client Side**

- 👍 Simple build & verification workflow
  - Open source tracker + Poplin extension
- 👍 Discover environment + browser capabilities
- 👍 No application load for interaction tracking

Benefits of **Server Side**

- 👍 Not impacted by ad-blockers
- 👍 Unaffected by Browser / client network errors
- 👍 No sending/leaking of (secret) context data
- 👍 No front end release required

Thank you!