



Canopy: An End-to-End Performance Tracing And Analysis System



Jonathan Kaldor¹
Bill Schaller¹

Jonathan Mace²
Pingjia Shan¹

Michał Bejda¹
Brendan Viscomi¹

Edison Gao¹
Vinod Venkataraman¹

Wiktor Kuropatwa¹
Kaushik Veeraraghavan¹

Joe O'Neill¹

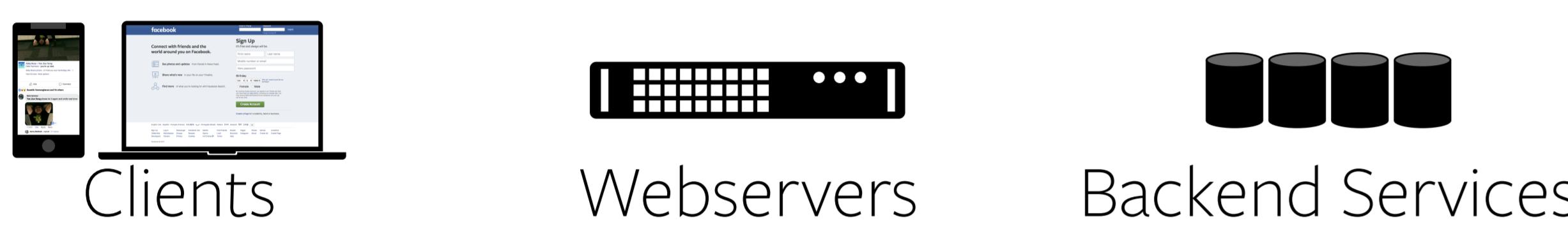
Kian Win Ong¹
Yee Jiun Song¹

¹Facebook

²Brown University

Goal

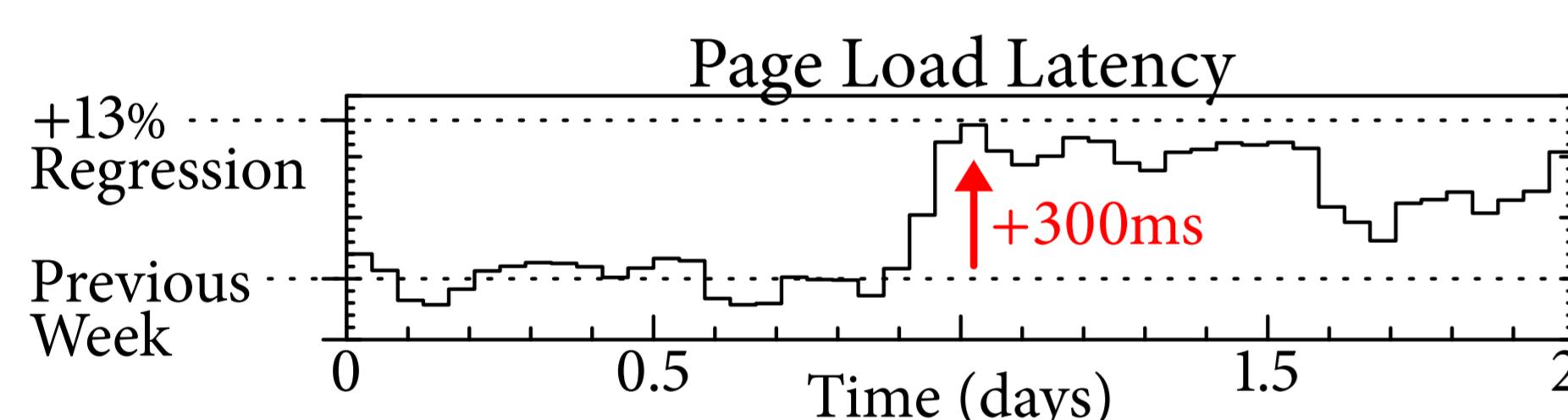
Diagnose performance problems that span many components



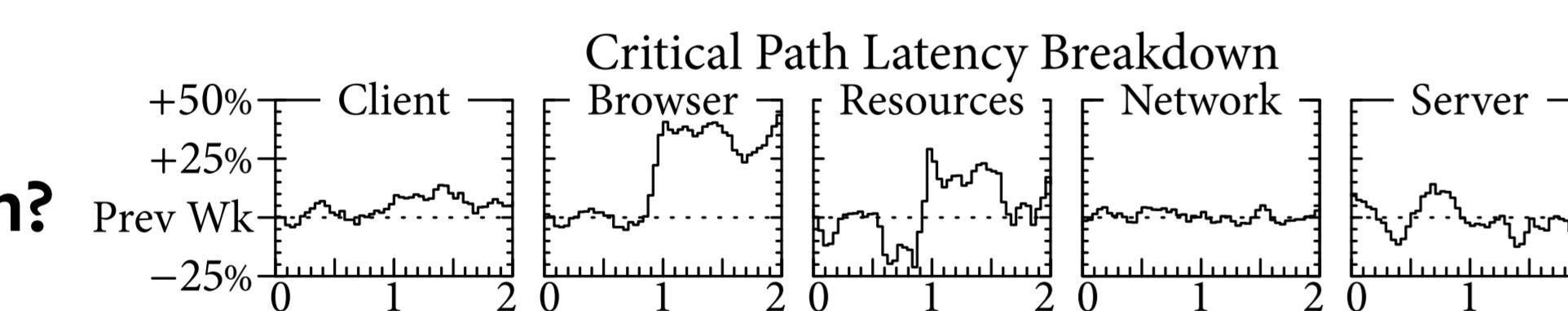
Support ad-hoc, exploratory analysis over large volumes of data

Gracefully support new and legacy components, new kinds of analysis

Average page load latency increased by 300ms



Where did the latency come from?



Why did resource loading increase?

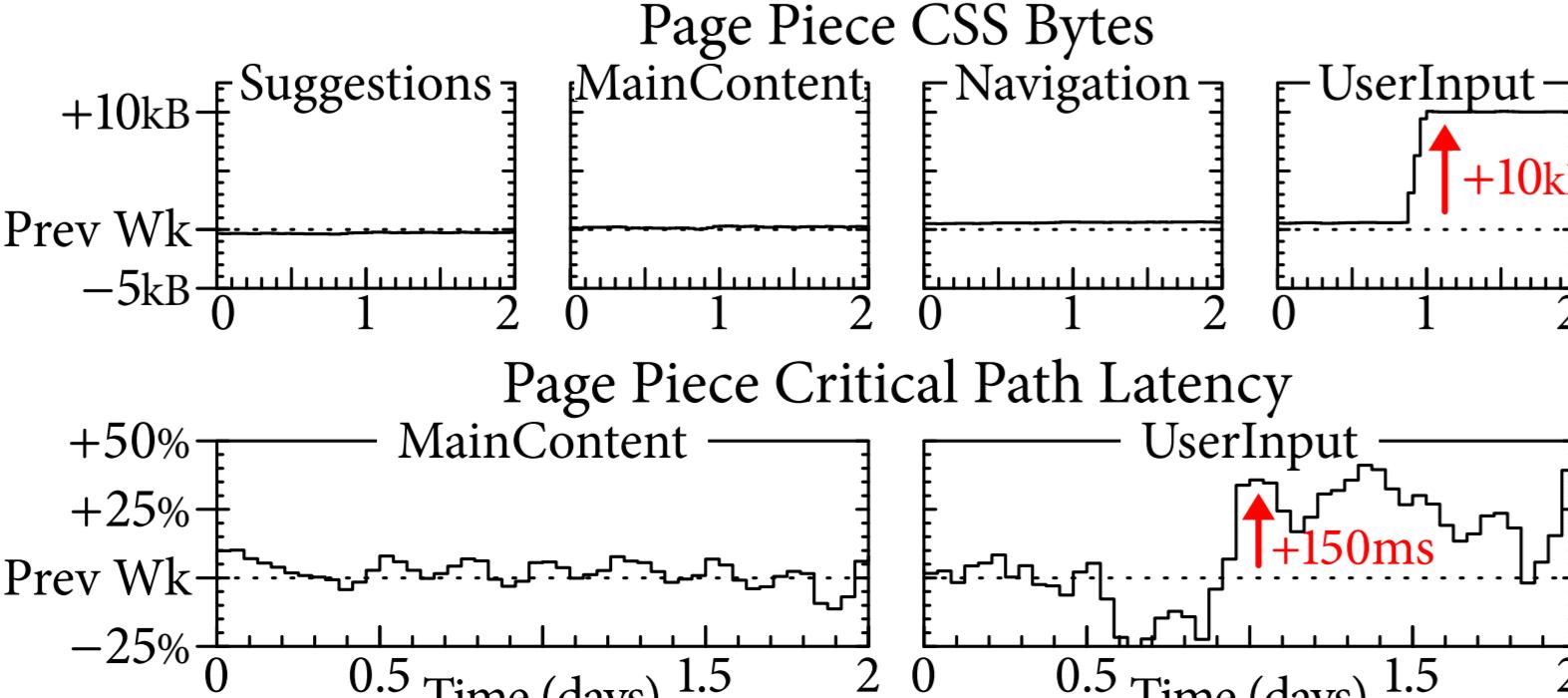
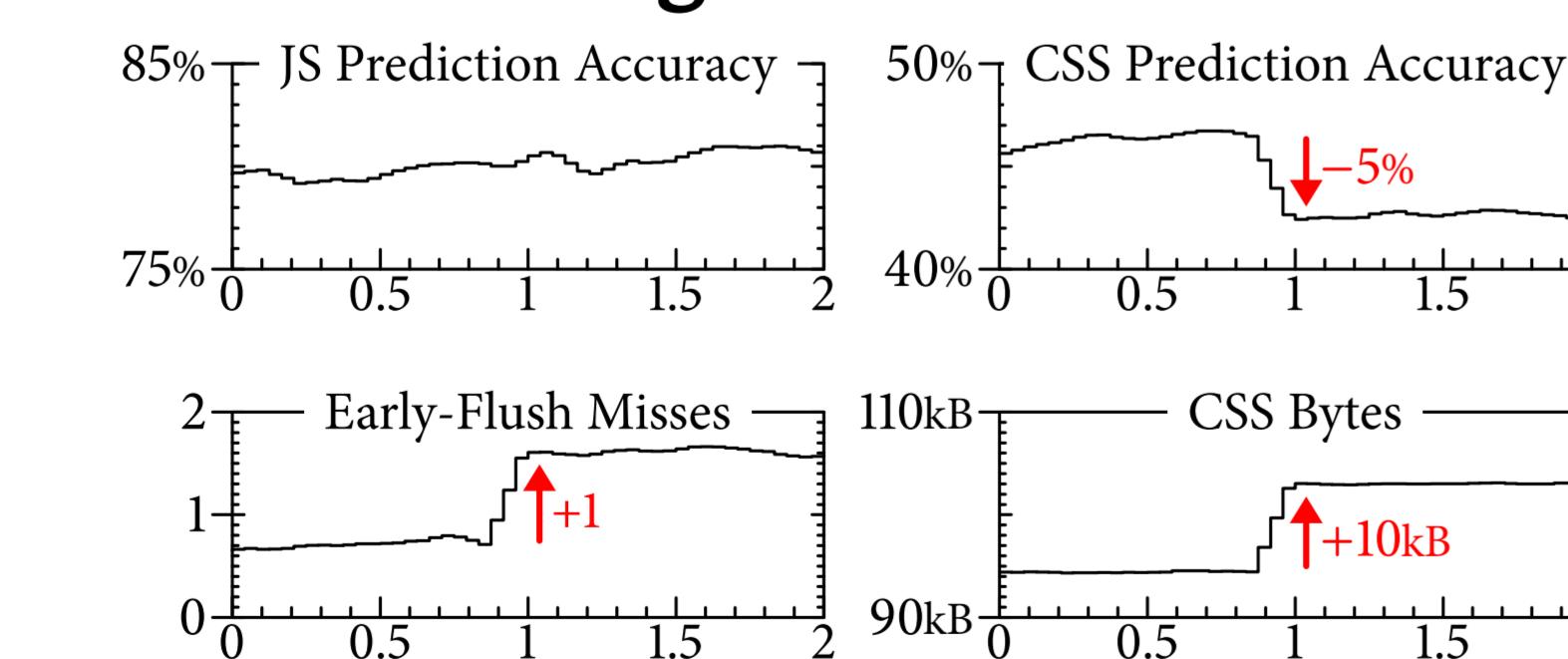
Worse prediction of required resources

Increase in CSS bytes before page display

Why did CSS prediction change?

UserInput needed an extra 10kB of CSS

UserInput blocked the page from displaying



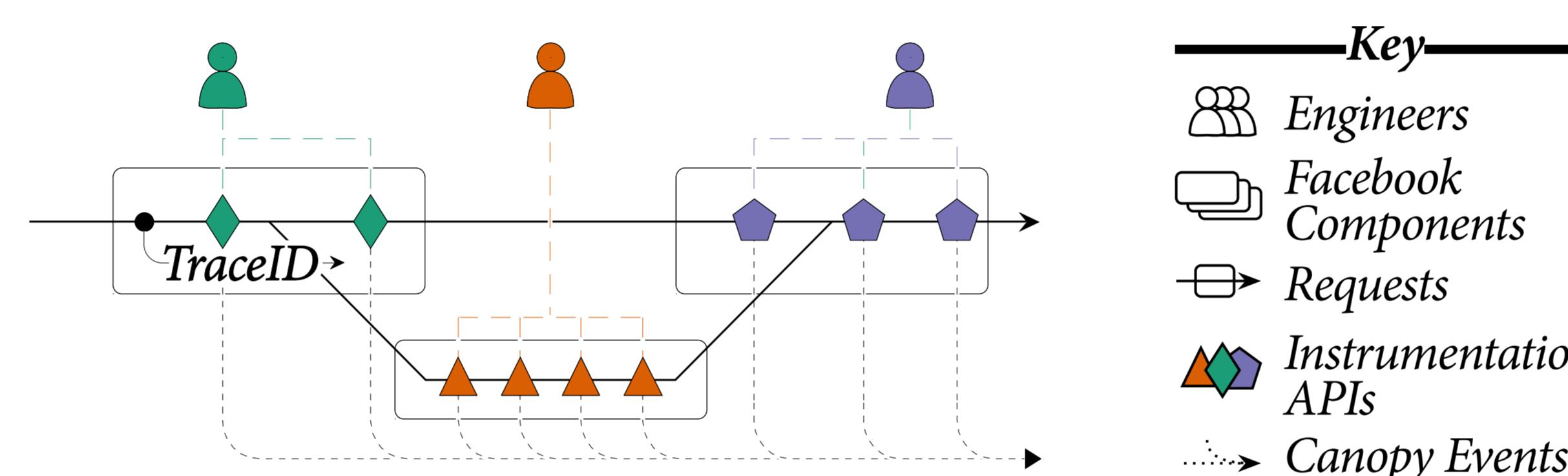
1

Instrumentation

Capture a wide range of performance data e.g., stack samples, counters, annotations, etc.

Agnostic to different execution models e.g., browser event loops, RPC calls, queues, async, CDN downloads

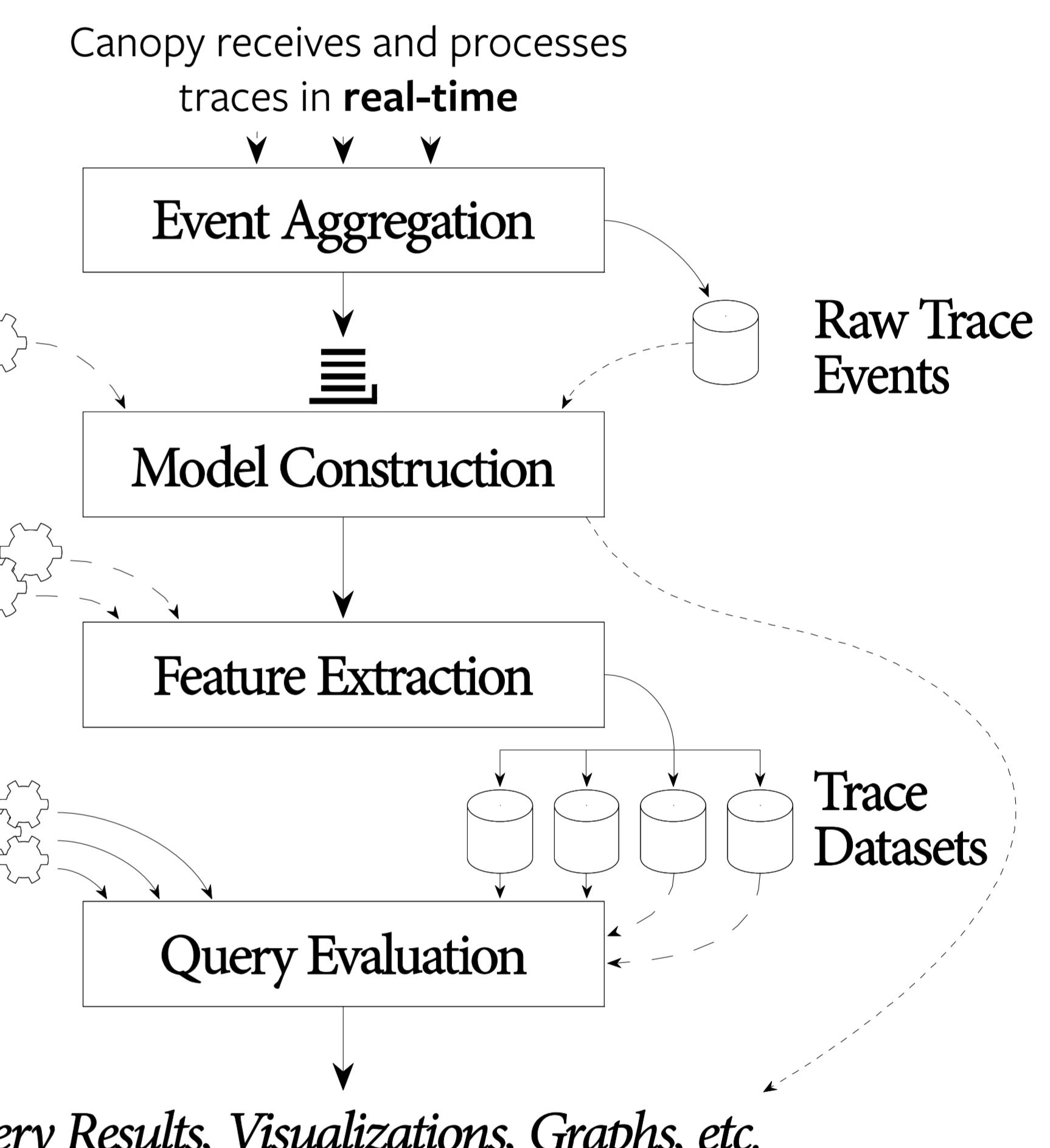
Propagate a **TraceID** at runtime to tie together trace events



2

Canopy Processing Pipeline

All events with the same TraceID route to the same backend shard



3

Feature Extraction

Users write **feature extraction functions** to extract **features** from modeled traces. A feature is a computed value or property.

Canopy provides a **domain-specific language** for extracting features as a pipeline of transformations, e.g.

DisplayDone = ExecUnits | Filter(name="Client") | Points | Filter(marker="display_done") | Timestamp

Canopy backends evaluate the feature extraction functions and pipe results to backend databases that can be interactively queried.

Each trace maps to a row and each feature maps to a column, e.g.,

| TraceID | Country | Page Load Latency | JS Prediction Accuracy |
|-------------|---------|-------------------|------------------------|
| EhjUmH4kaPG | USA | 1000ms | 85% |

Canopy users can pipe results to their own output tables and apply custom feature extraction functions.

4

Summary

- Canopy processes **> 1 billion traces per day**
- Canopy has been deployed in production for the past **2 years**
- Canopy imposes **<10%** runtime overhead to trace a request
- Canopy further amortizes runtime costs by **sampling** requests
- Canopy's backend gracefully handles load spikes and provides per-user resource isolation