



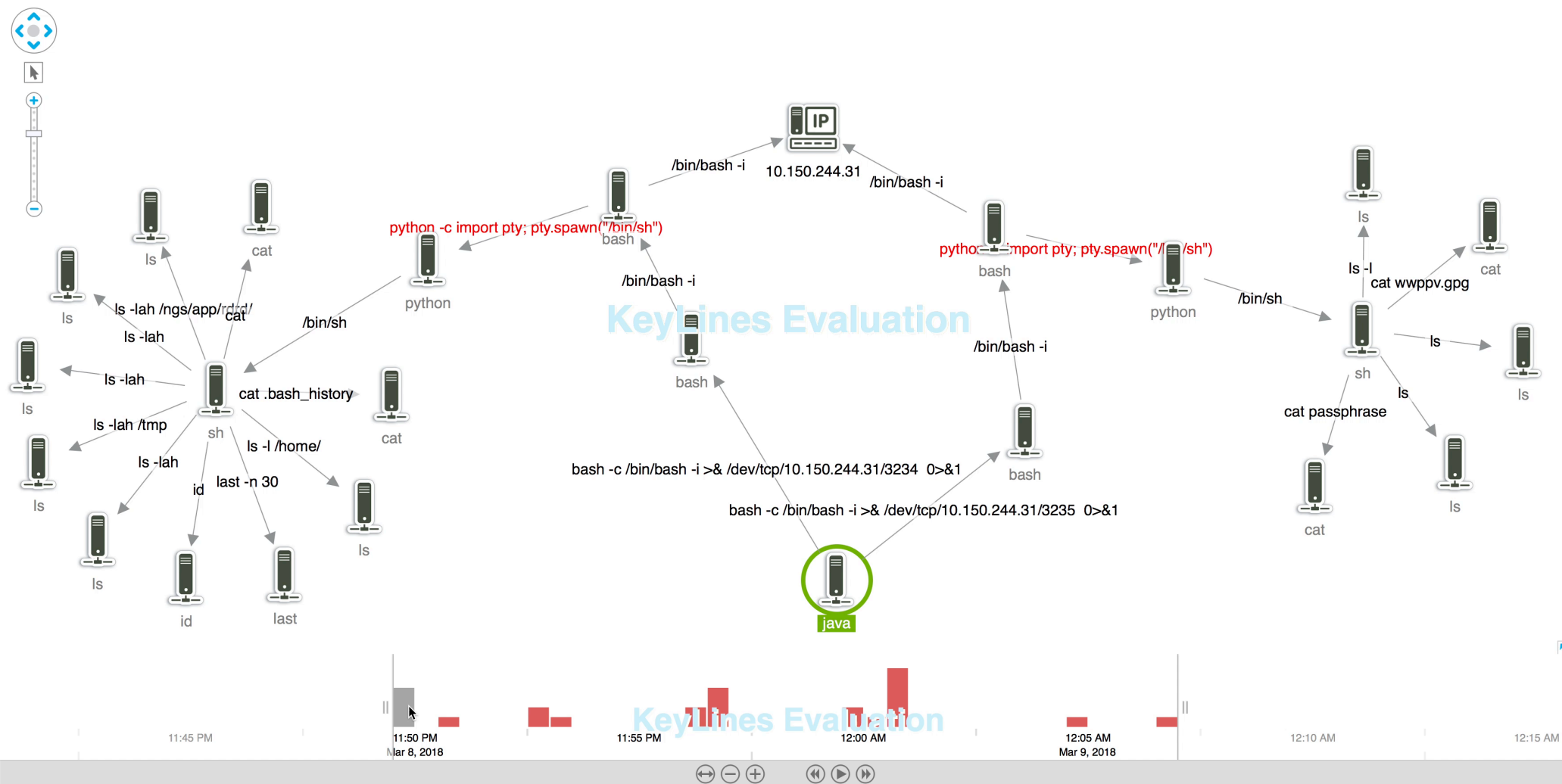
Threat Detection and Response at Scale

Dominique Brezinski — Apple Information Security

This is about the data platform aspect,
not the specific analytics

Agenda

Use Cases, Scale,
and Challenges/Solutions



Enabling Detection and Analytics

Diverse threats require diverse data sets

Streams (left joined) with context and
filtered or (inner joined) with indicators

Large time window, multi-dataset graphs

Enabling Triage and Containment

Search and Query

```
WHERE date > current_date() - 30 days
```

Scale

3.6m rec/s

Input rate

3.4m rec/s

Processing rate

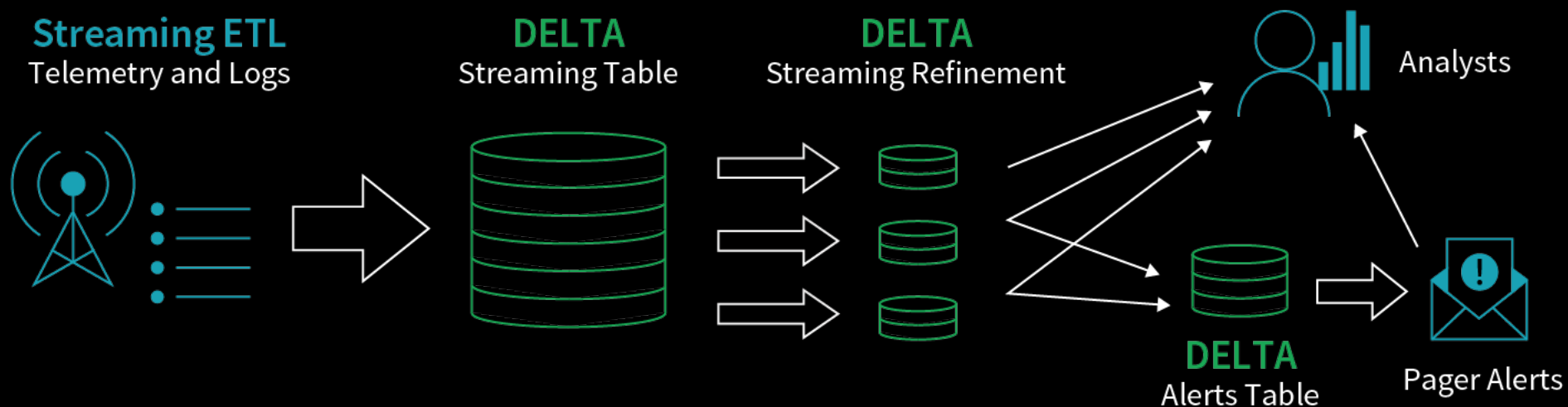
>100TB new data a day

>300 billion events per day

Most queried table:
504,761,911,529,518 bytes,
11,149,012,553,409 rows

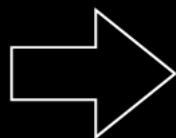
Yeah, trillions!

Streaming Ingestion Architecture



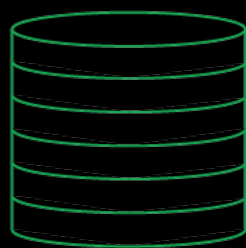
Streaming ETL

Telemetry and Logs



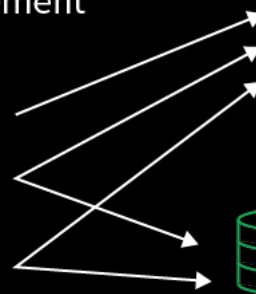
DELTA

Streaming Table



DELTA

Streaming Refinement



Analysts

DELTA

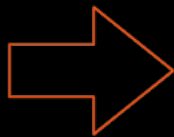
Alerts Table



Pager Alerts

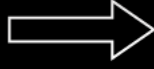
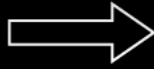
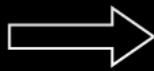
Streaming ETL

Telemetry and Logs



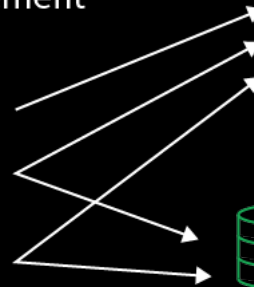
DELTA

Streaming Table



DELTA

Streaming Refinement



Analysts

DELTA

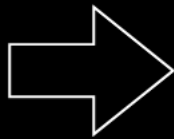
Alerts Table



Pager Alerts

Streaming ETL

Telemetry and Logs



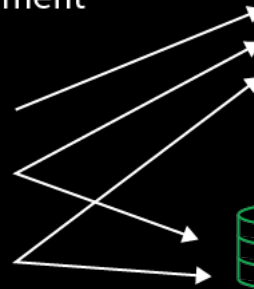
DELTA

Streaming Table



DELTA

Streaming Refinement



DELTA

Alerts Table



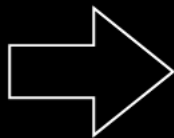
Analysts



Pager Alerts

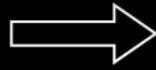
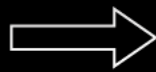
Streaming ETL

Telemetry and Logs



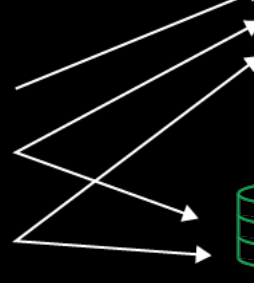
DELTA

Streaming Table



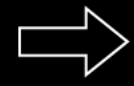
DELTA

Streaming Refinement



DELTA

Alerts Table



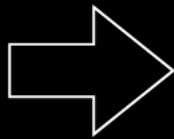
Analysts



Pager Alerts

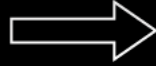
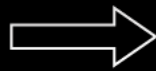
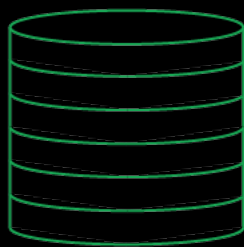
Streaming ETL

Telemetry and Logs



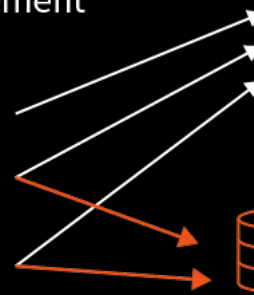
DELTA

Streaming Table



DELTA

Streaming Refinement



Analysts

DELTA

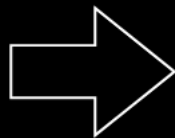
Alerts Table



Pager Alerts

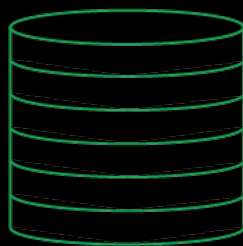
Streaming ETL

Telemetry and Logs



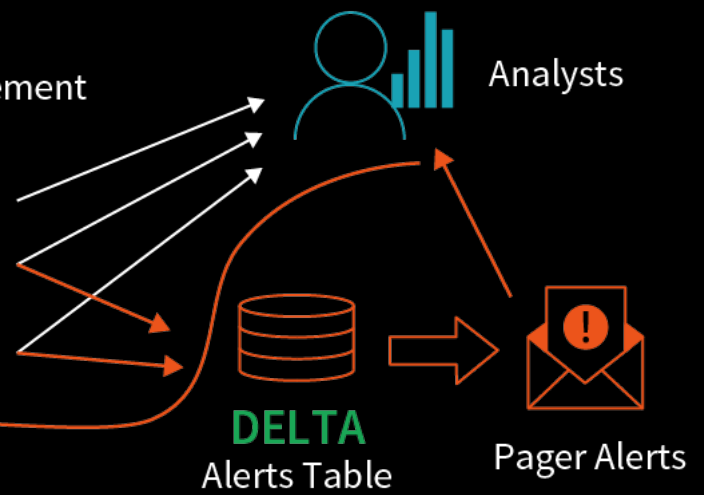
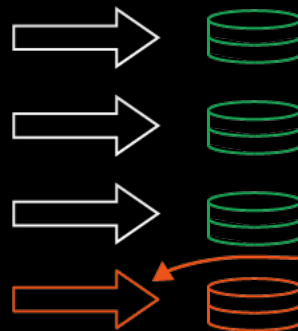
DELTA

Streaming Table



DELTA

Streaming Refinement



DELTA

Alerts Table

Analysts



Pager Alerts

WHERE src_ip = x AND dst_ip = y

Total data size: 504 terabytes, 11,149,387,374,965 rows

Scanned data size: 36.5 terabytes, 722,630,063,648 rows

Additional reduction thanks to data skipping (bytes): 92.4%

Additional reduction thanks to data skipping (rows): 93.2%

Simple. Unified.