



Can You Roll Your Own SIEM?

Ethan Christ & Bret Rubin



Introduction

Bret Rubin

Security Engineer, Security Infrastructure

Ethan Christ

VP, Security Identity, Monitoring and Response

Two Sigma is a Financial Sciences company that brings a scientific approach to financial services. We do that by applying data science, technology and systematic research to investment management, private equity, venture capital, impact investing, insurance technology and market making.

<https://careers.twosigma.com/>



Important Legal Information

This document is being distributed for informational and educational purposes only and is not an offer to sell or the solicitation of an offer to buy any securities or other instruments. The information contained herein is not intended to provide, and should not be relied upon for, investment advice. The views expressed herein are not necessarily the views of Two Sigma Investments, LP or any of its affiliates (collectively, “Two Sigma”). Such views reflect the assumptions of the author(s) of the document and are subject to change without notice. The document may employ data derived from third-party sources. No representation is made by Two Sigma as to the accuracy of such information and the use of such information in no way implies an endorsement of the source of such information or its validity.

The copyrights and/or trademarks in some of the images, logos or other material used herein may be owned by entities other than Two Sigma. If so, such copyrights and/or trademarks are most likely owned by the entity that created the material and are used purely for identification and comment as fair use under international copyright and/or trademark laws. Use of such image, copyright or trademark does not imply any association with such organization (or endorsement of such organization) by Two Sigma, nor vice versa



Agenda

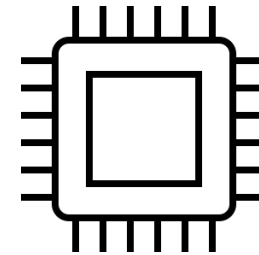
- Considerations & Requirements
- Decision Rationale
- Technical Implementation and Details
- Demo
- Tactical Results
- Strategic Implications



Considerations & Requirements

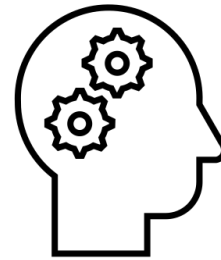
Considerations & Requirements

- Quantity
 - 2+ TB/day with room to grow
- Speed
 - <60s from event to searchable



Considerations & Requirements

- Reliability
 - Forensics & availability
- Flexible data parsing and ingestion logic
 - Bespoke logging standards



Considerations & Requirements



- Retention
 - Up to ∞
- Searchability
 - Speed of large range queries
 - User friendliness

Considerations & Requirements



- Security
 - Sensitive data and ACLs
- Alerting
 - Automated downstream actions
 - Multiple notification paths

Noted Omissions



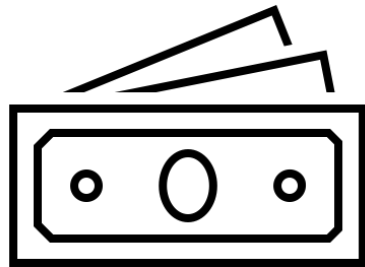
Threat Model





Decision Rationale

Build vs. Buy



- Upfront Cost
 - Licensing
 - Infrastructure
 - Business drivers

Build vs. Buy


- Pre-existing Investment
 - BQ Slots
 - Storage rate
 - Established Interconnect





Feature Parity

How does what comes with a vendor SIEM product compare to what we'd have to write ourselves, excluding equivalent setup work?



Implementation Details - Ingestion



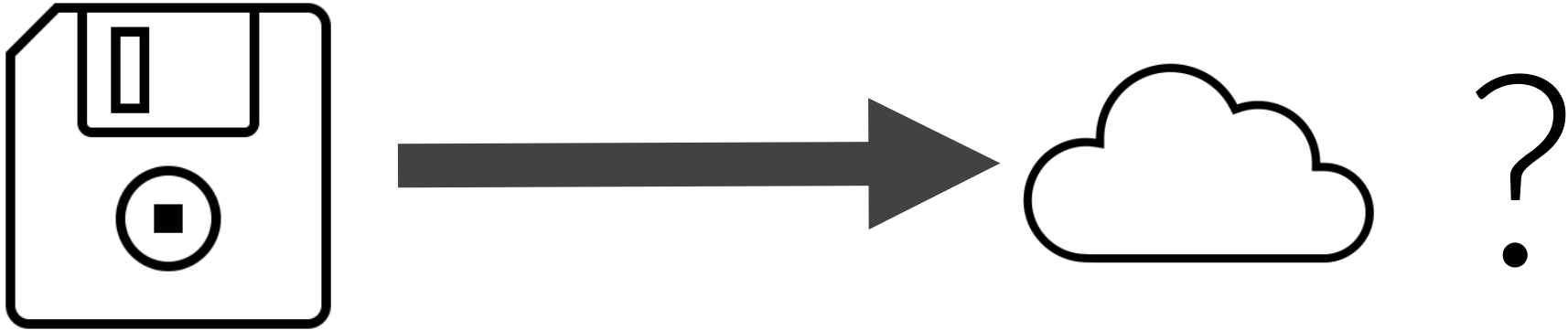
What We Needed

- ❑ batch file loads from system appliances
- ❑ streaming ingest from configurable sources
- ❑ scheduled query execution for complex logic
- ❑ streaming alerting for single-line pattern matches

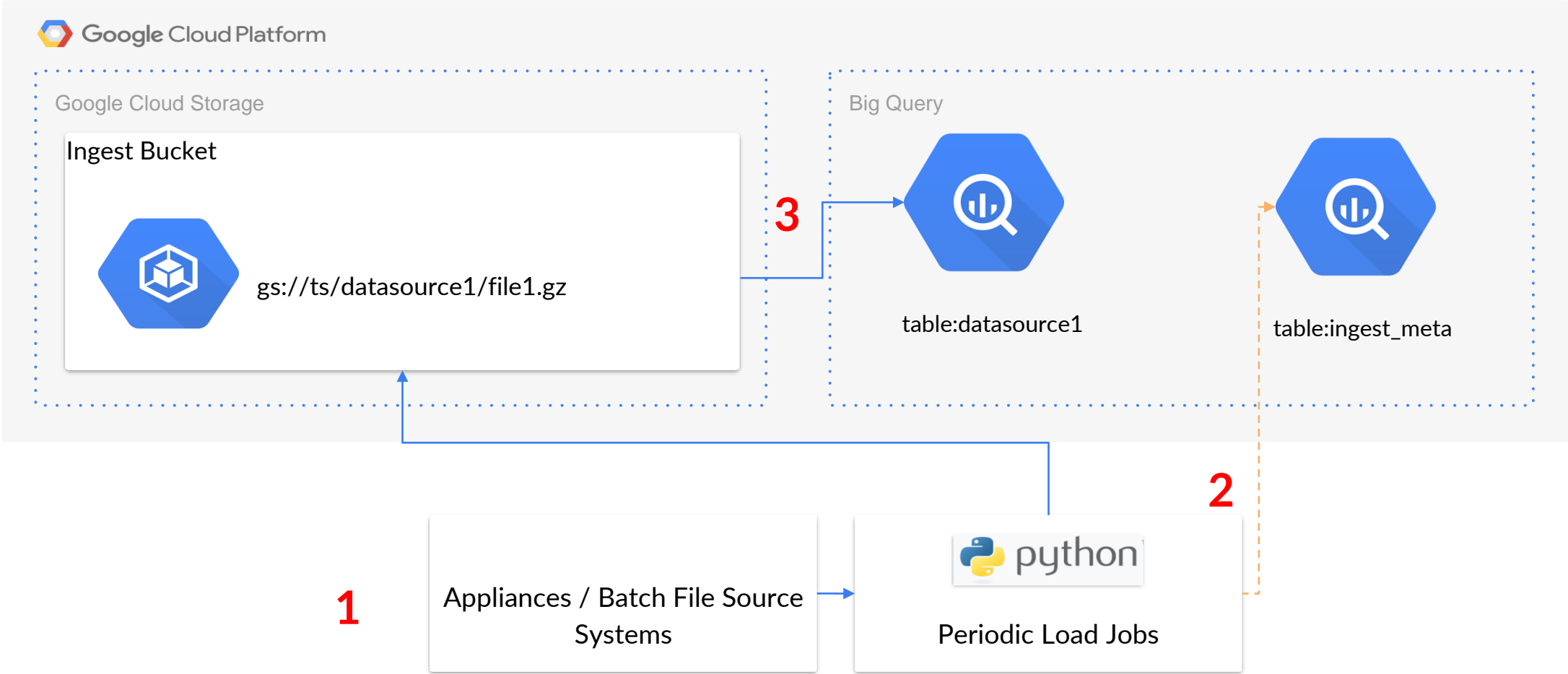
bonus points...

- ❑ revision control / peer review for scheduled query definitions
- ❑ alert output routing to email, jira, pagerduty, slack...

Batch Loads



Batch Loads

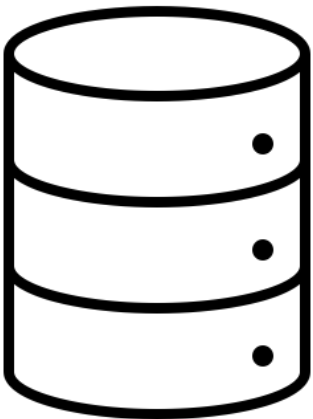


Streaming Ingest



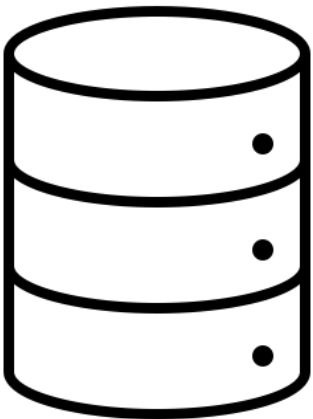
Streaming Ingest

- reliable log forwarding from linux base platform (fluentbit)
 - route directly to GCP via interconnect
 - native load balancing via GCP APIs
 - fluentbit efficient and fast



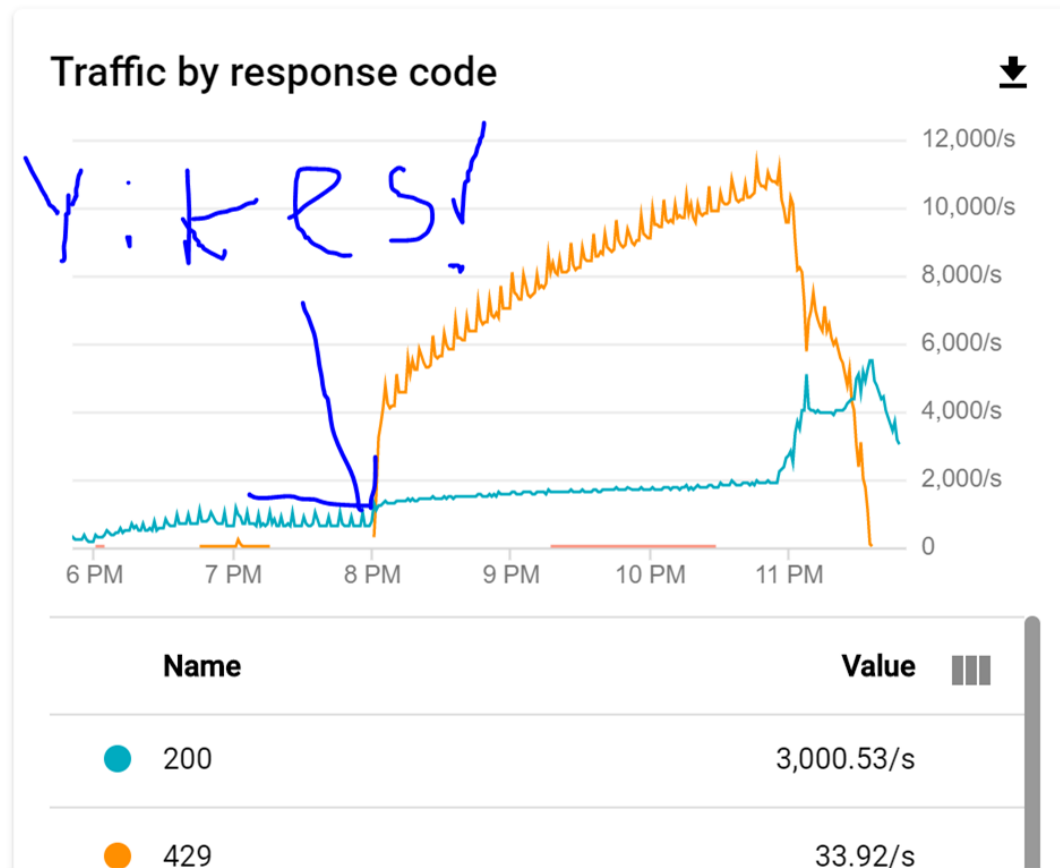
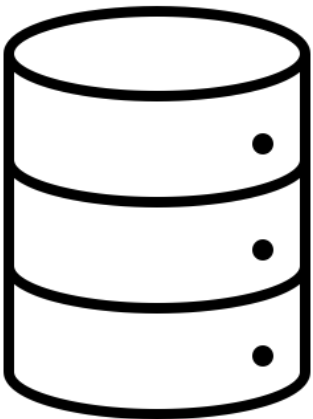
Streaming Ingest

- complex parsing, routing, enrichment (fluentd aggregator instances)
 - receiver for logs from bespoke systems
 - appliances in DMZ and other limited connectivity environments
 - fluentd allows flexible and performant parsing and routing



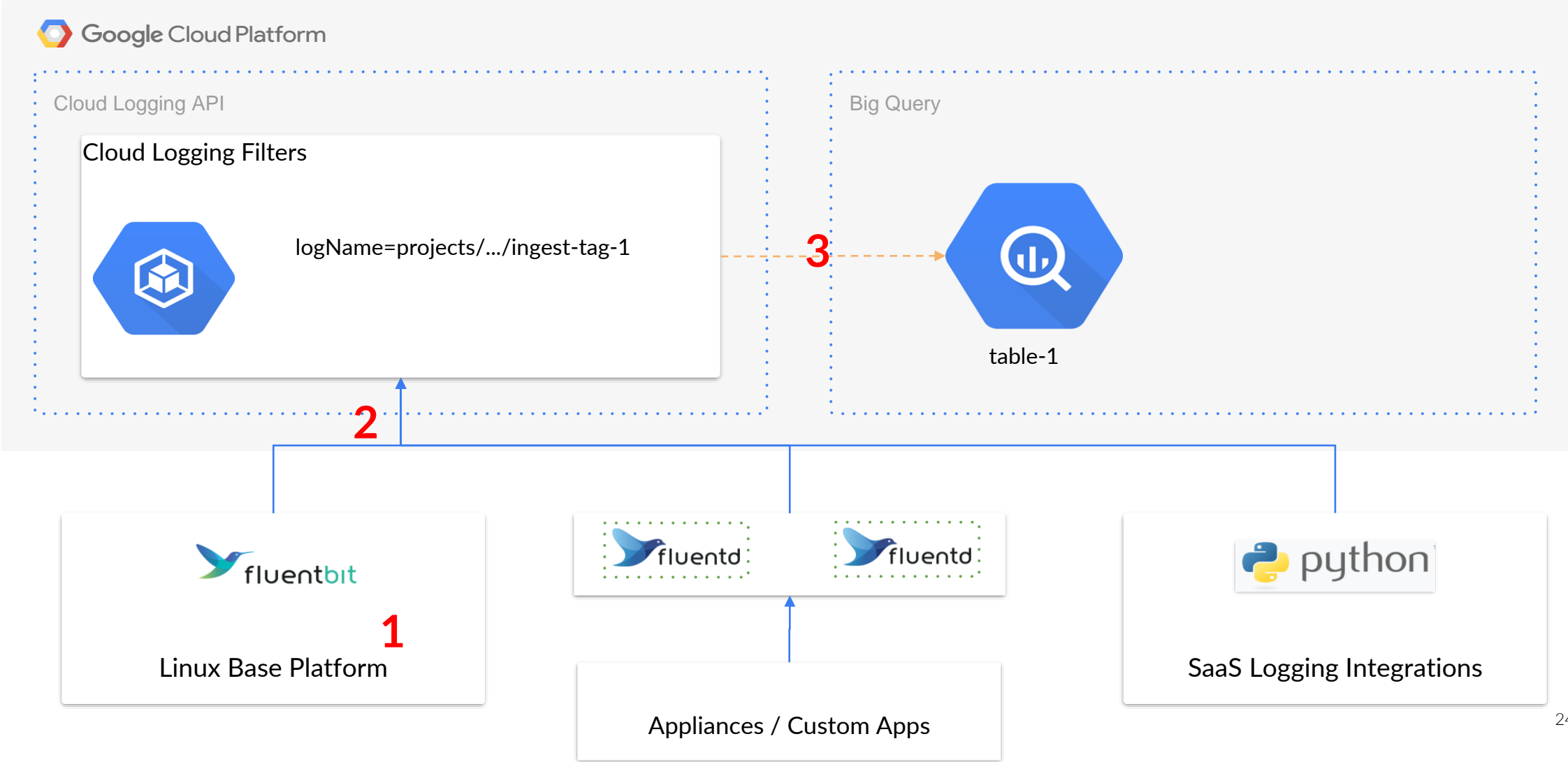
Streaming Ingest

- SaaS service
 - vendor help
 - internal to admins
 - goal to migrate team



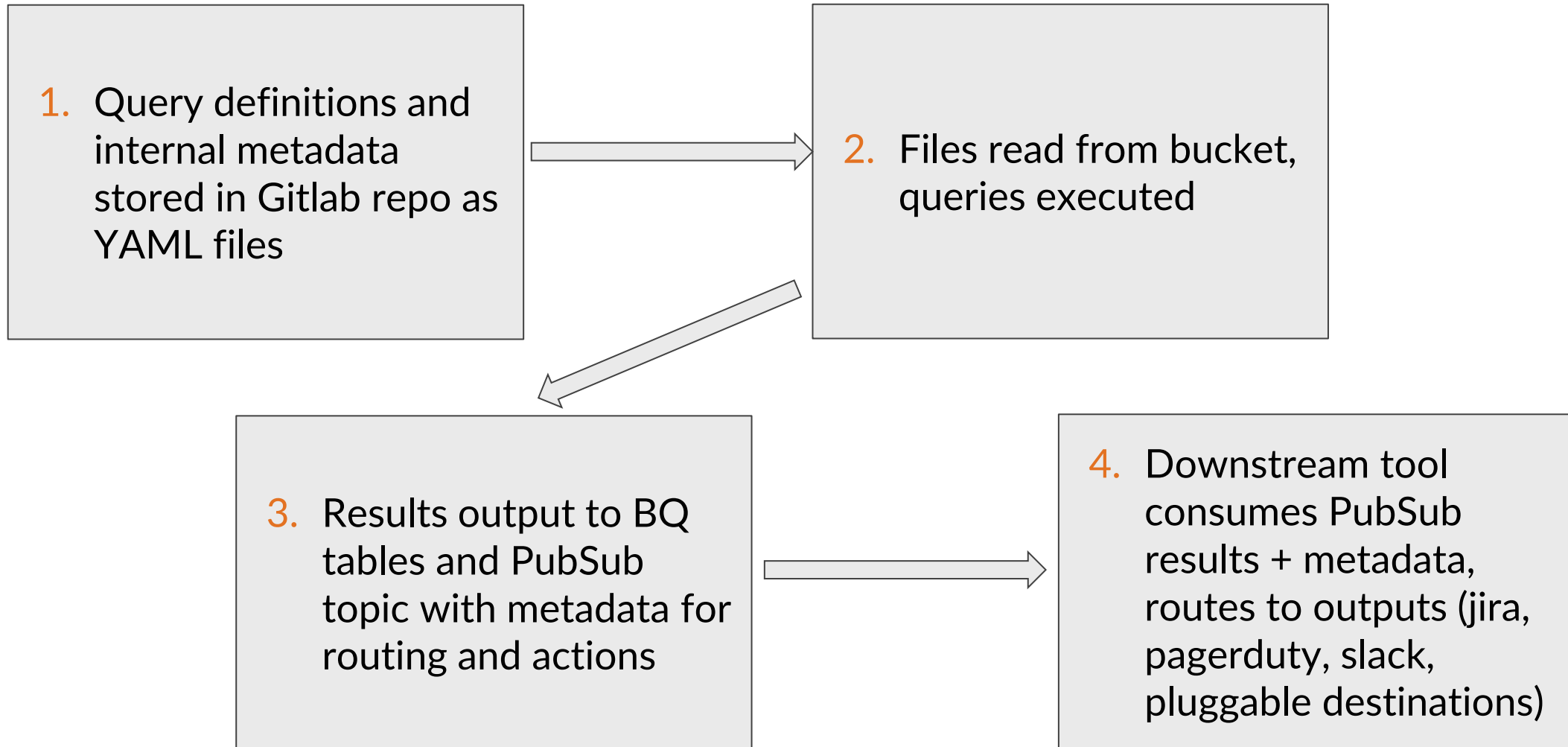
pushed to API
opers and
ay from security

Streaming ingest

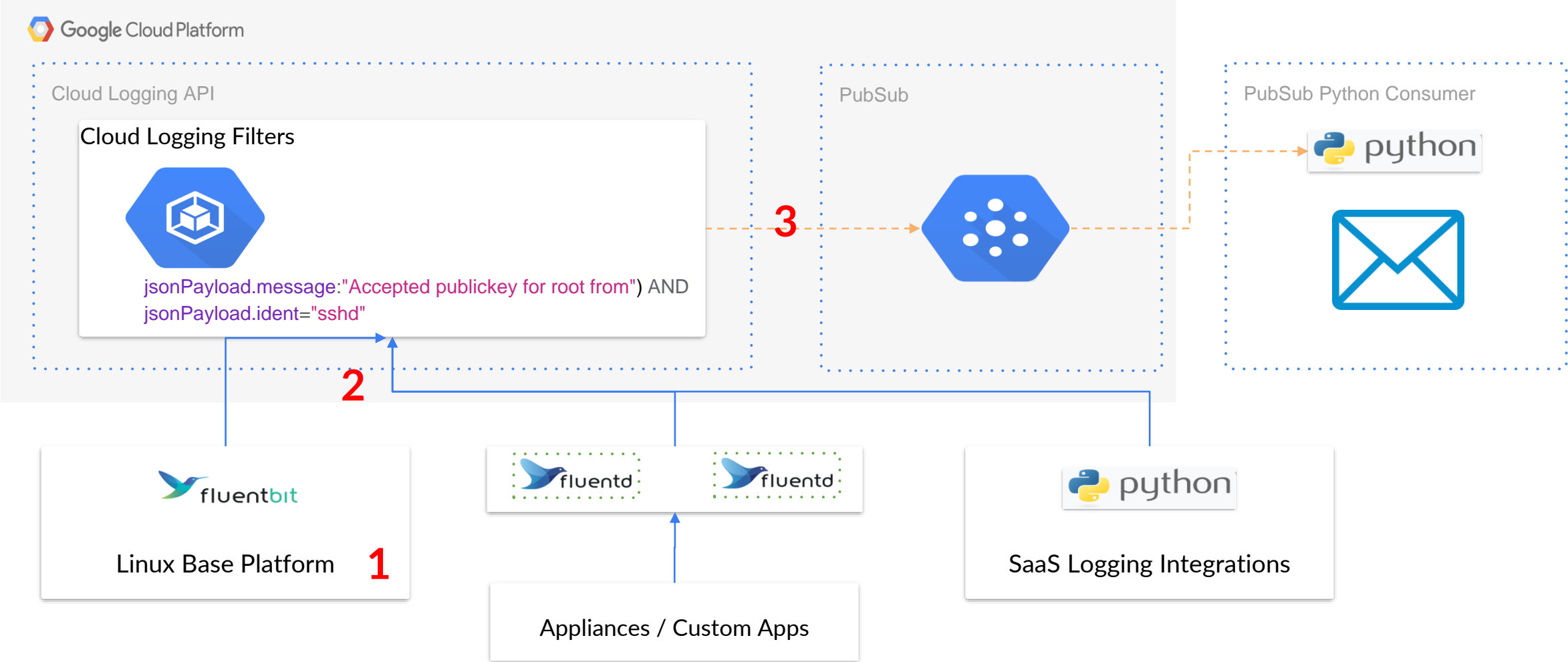


Queries & Alerting

Scheduled Queries



Streaming Alerting





Access Controls



Data Access Controls

- per-dataset ACLs synced from internal directory service
- multi-party signoffs for access
- audit logs of user and automation queries



Operational Wins



Operational Wins

- **no mores**
 - no more “log rehydration” (longer retention / no query slowdown)
 - no more “too big for our SIEM” volume cost / security value tradeoffs
- **high fives**
 - high five: query over years / petabytes of data in “reasonable interactive time”
 - high five: development speed / offload of some data pipeline management



Lessons Learned



Lessons Learned

- **cloud service offerings change fast**
- **in fast moving environments, monitoring is critical**
- **streaming costs are slight, but they do add up!**



Demo



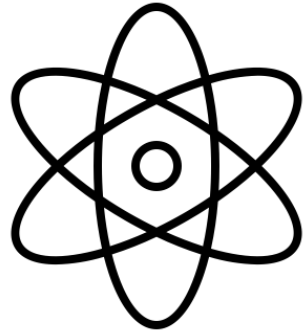
Tactical Results

Overall Effort

- 6,000 lines of code
- ~9 months FTE time



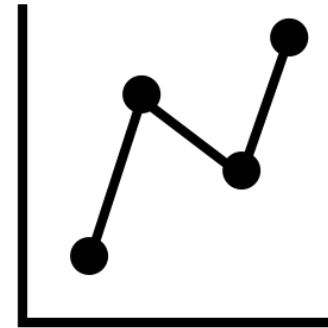
Resultant Capabilities



- 5 PB stored today
- > 5TB/day ingestion
- Improved Performance

Cost Savings

- ~\$3.5M in licenses
- \$600k in annual maintenance
- .1 FTE/year in staffing resources





Strategic Implications

Data, Data, Data

- New data feeds
- Open experimentation

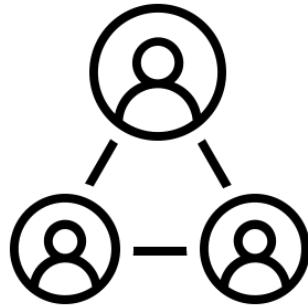


Search & Alerting



- Staff training
- Detection and response

Observability & Flexibility



- Partnerships with engineering teams
- Extensible parsing and data integration



What's Next?

Conclusion

Questions?