

# Project Shield

How we use Beam to defend democracy and free expression, and how we got started!

Marc Howard  
Staff Engineer - Google

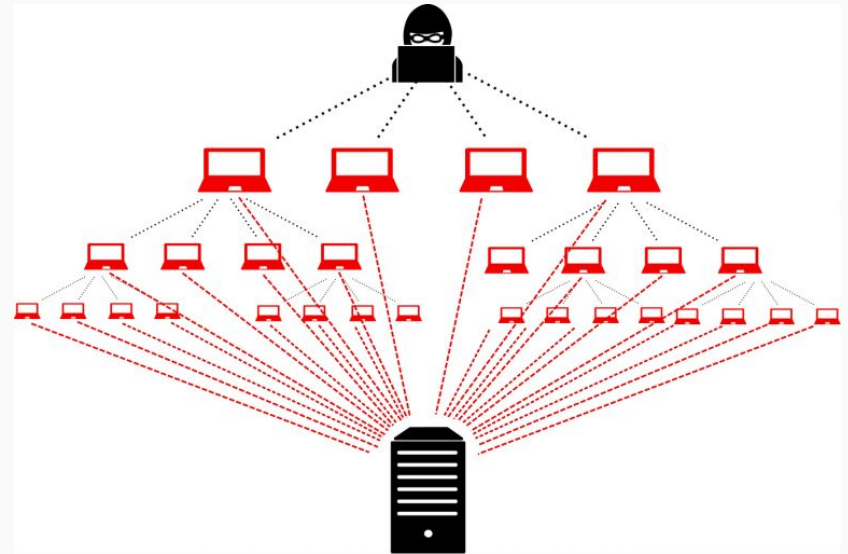


BEAM  
SUMMIT

September 4-5, 2024  
Sunnyvale, CA. USA

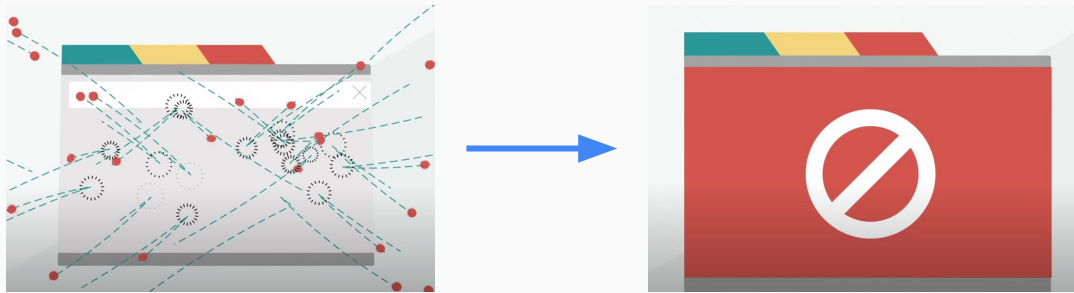
# DDoS Attacks

- **Distributed Denial of Service**
- Common attack to take down websites
- Requires no special access



# Self-Defense is Hard

- Huge volume of traffic overloads server
- Cannot defend against alone





# Impact of DDoS



SECURITY

## Cyberattacks Rise as Ukraine Crisis Spills to Internet

By NICOLE PERLROTH MARCH 4, 2014 6:47 PM 4 Comments

Email

Share

Tweet

Save

More



The crisis in Ukraine has spread to the Internet, where hackers from both sides are launching large cyberattacks against opposing news organizations.

Security experts say that they are currently witnessing unusually large denial-of-service attacks, also called DDoS attacks, in which hackers flood a website with traffic to knock it offline. The attacks have been directed at both pro-Western and pro-Russian Ukrainian news sites.

In at least one case, hackers successfully defaced the website of the Kremlin-financed news network Russia Today, replacing headlines and articles containing the word "Russia" with the word "nazi."

Experts say the attacks on pro-Western Ukrainian news sites closely

**THE VERGE** TRENDING NOW Tesla added a feature back to the Model S after Elon Musk's kid called it 'the stupidest car in...

LOG IN | SIGN UP LONGFORM VIDEO REVIEWS TECH SCIENCE ENTERTAINMENT DESIGN BUSINESS US & WORLD FORUMS

PREVIOUS STORY Howard the Duck has officially made a comeback

NEXT STORY Microsoft is bringing chat to its Word and PowerPoint web apps

US & WORLD **7** COMMENTS

## Before protests, Hong Kong news sites were hit with the largest DDoS ever

By Russell Brandom on November 21, 2014 05:59 pm Email @russellbrandom

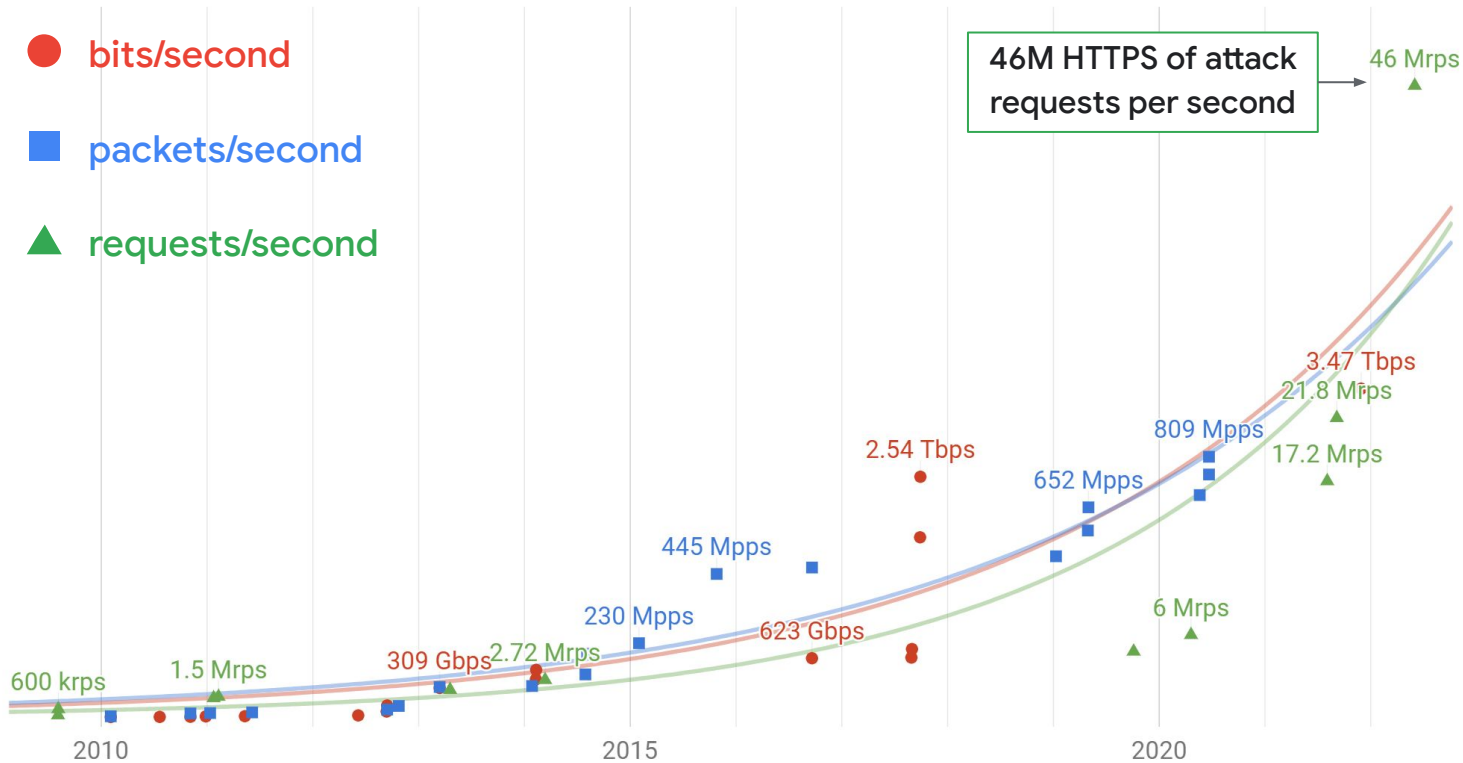
DON'T MISS STORIES FOLLOW THE VERGE Like Follow Subscribe Follow

SIGN UP FOR OUR NEWSLETTER



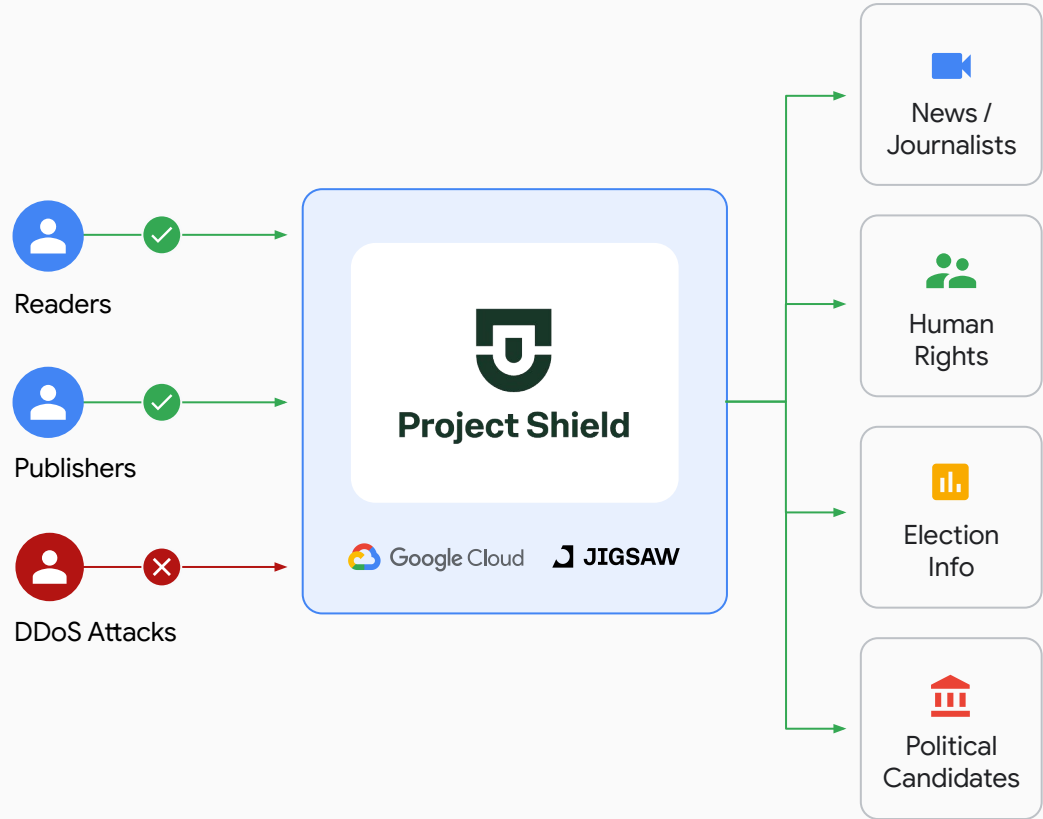
BEAM  
SUMMIT

# Growing in Size

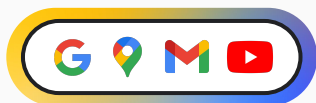
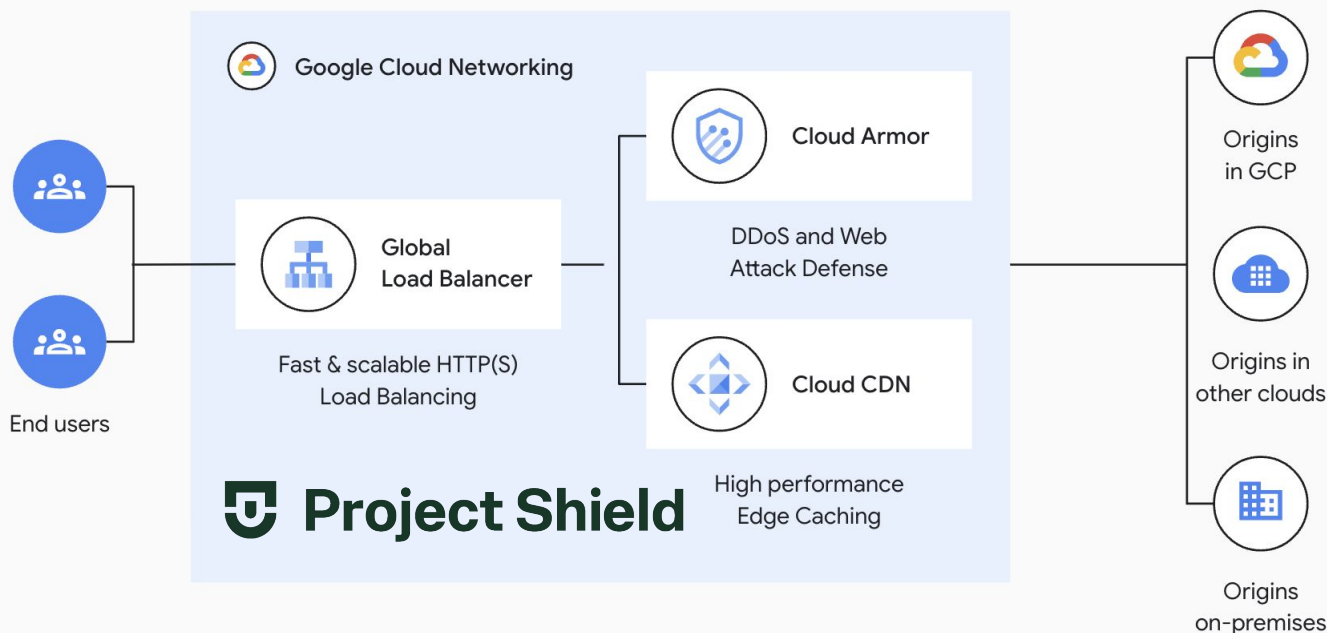


# Project Shield

- Support elections
- Protect information
- Free defense



# Built on Cloud Networking



Same technologies, infrastructure, and teams that deliver, scale and protect Google apps with billions of users



BEAM  
SUMMIT



# Built on Cloud Networking

## One of the largest networks in the world

40 cloud regions (+10 announced)

121 zones

187 network edge locations

29 subsea cables

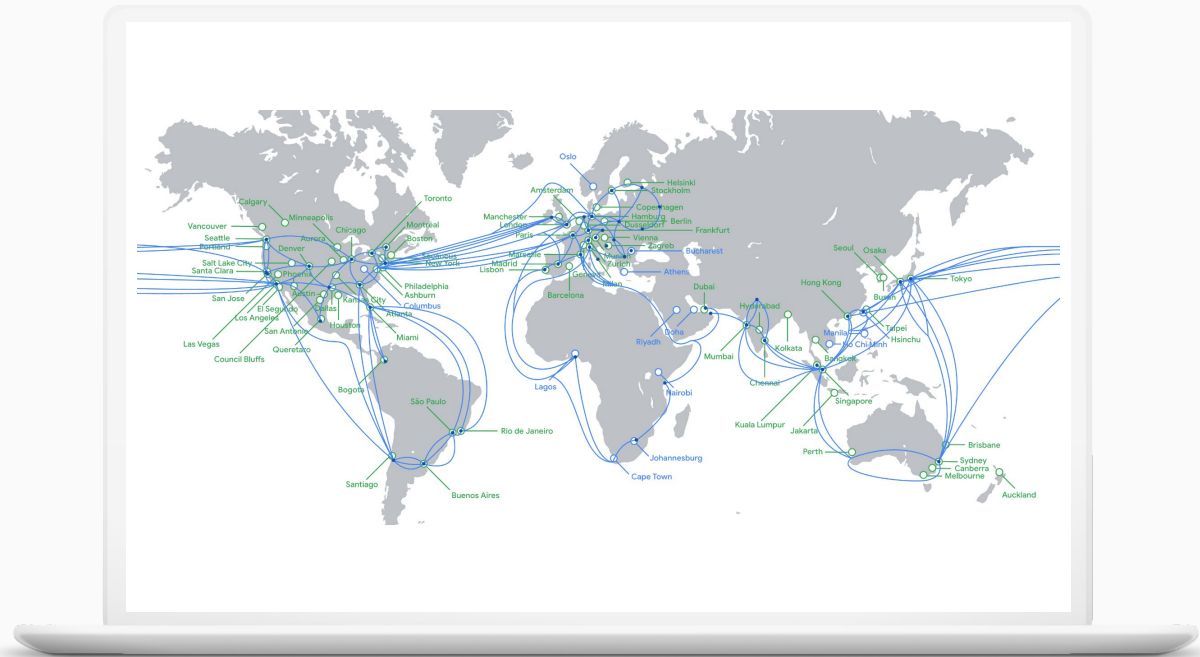
200+ countries and territories

100+ interconnect locations

100%

Renewable on an annual basis since 2017

<https://cloud.google.com/about/locations>





# Defense Results

- 99.9% DDoS attack traffic volume mitigated
- >99.99% uptime for DDoS Attack victims
- Continuous attack identification and real-time defenses tuning for best results.



## Project Shield



BEAM  
SUMMIT

# Impact of Shield



## Project Shield

Mykhailo Fedorov presented the first Ukraine Peace Prize to Google

Google has also expanded cybersecurity for Ukrainian accounts. Project Shield has become very important for our state resources and media - Ukraine is grateful for free unlimited protection against DDoS attacks.

**KrebsOnSecurity**  
In-depth security news and investigation

HOME ABOUT THE AUTHOR ADVERTISING/SPEAKING

How Google Took on Mirai, KrebsOnSecurity



**BEAM**  
SUMMIT

WIRED BACKCHANNEL BUSINESS CULTURE GEAR IDEAS MORE v SIGN IN SUBSCRIBE

ANDY GREENBERG SECURITY FEB 24, 2016 10:00 AM

### Google Wants to Save News Sites From Cyberattacks—For Free

It hopes to extend its so-called Project Shield DDoS attack protection to tens of thousands of websites.

GOOGLE NEWS INITIATIVE

### Project Shield: Defending Maka Angola

Security & Identity

### How Project Shield helped protect U.S. midterm elections from DDoS attacks

Security & Identity

### Google mitigated the largest DDoS attack to date, peaking above 398 million rps

October 10, 2023

### Google CEO Sundar Pichai announces Project Shield in first visit to Europe

In his first public appearance in Europe after becoming Google's CEO, Sundar Pichai made important product announcements aimed at publishers, including the launch of Project Shield

# Data-driven Defense

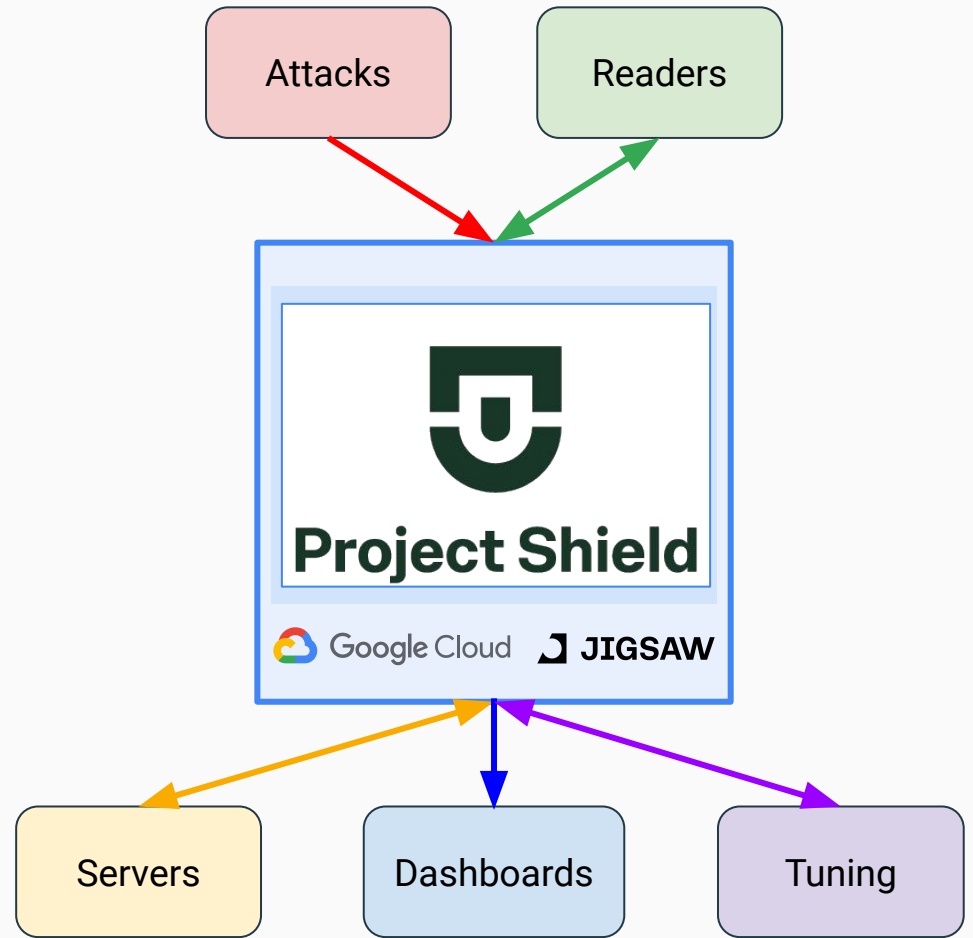
Shield handles many types of data

Traffic is served and analyzed

- Served from cache
- Rate limited
- Sent to hosting server

Users are shown metrics

Defenses are improved





# Logs at Uneven Scale



Project Shield defended 4x the usual number of attacks during Oct-Nov'22.

The attacks were against political parties and elections-related news.



# Costly Analysis

- Redundant calculation
- Poor serving latency

DDoS Traffic Logs

Many Use Cases

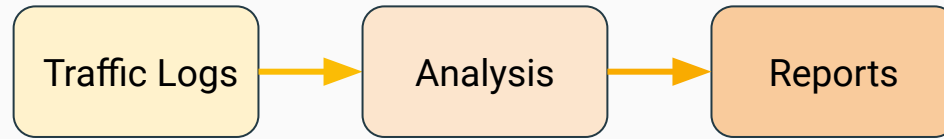
Repeated Lookups

Expensive

Slow

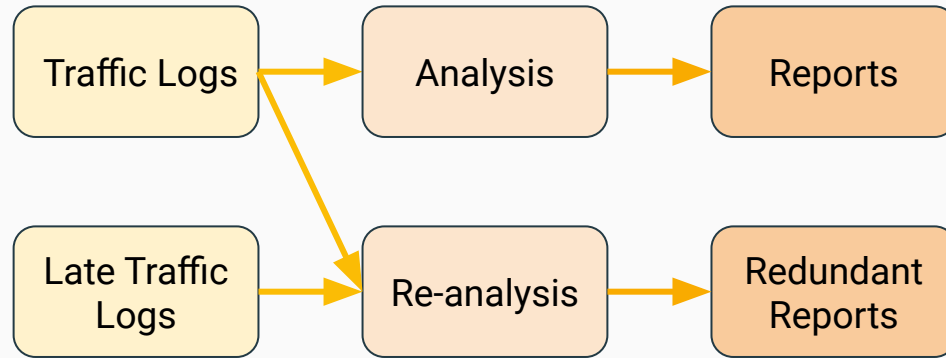


# Bespoke Preprocessing

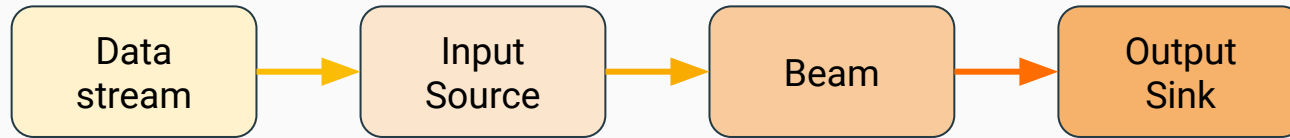




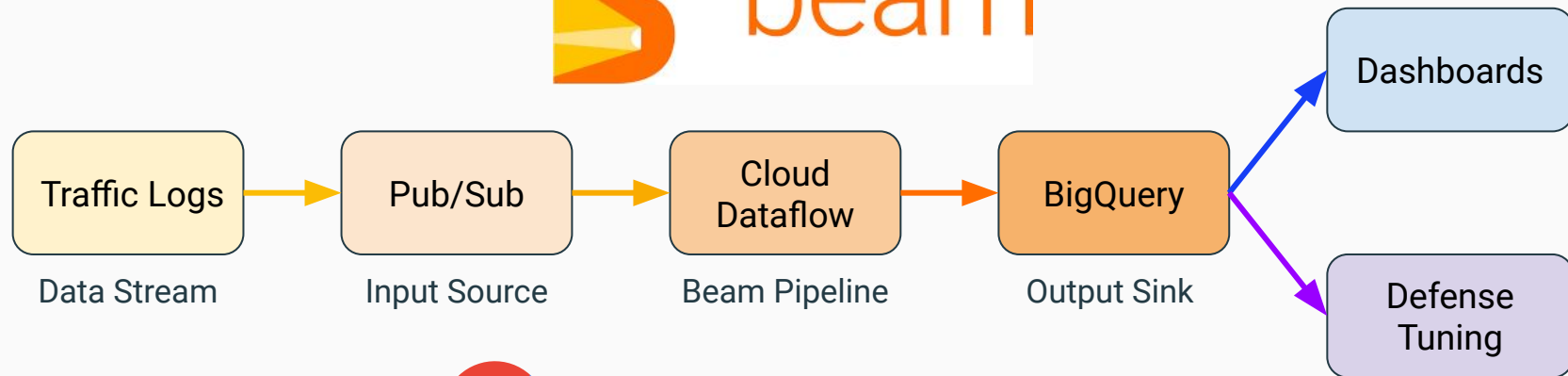
# Falling behind



# Data Pipelines



# Beam on Google Cloud



Google Cloud

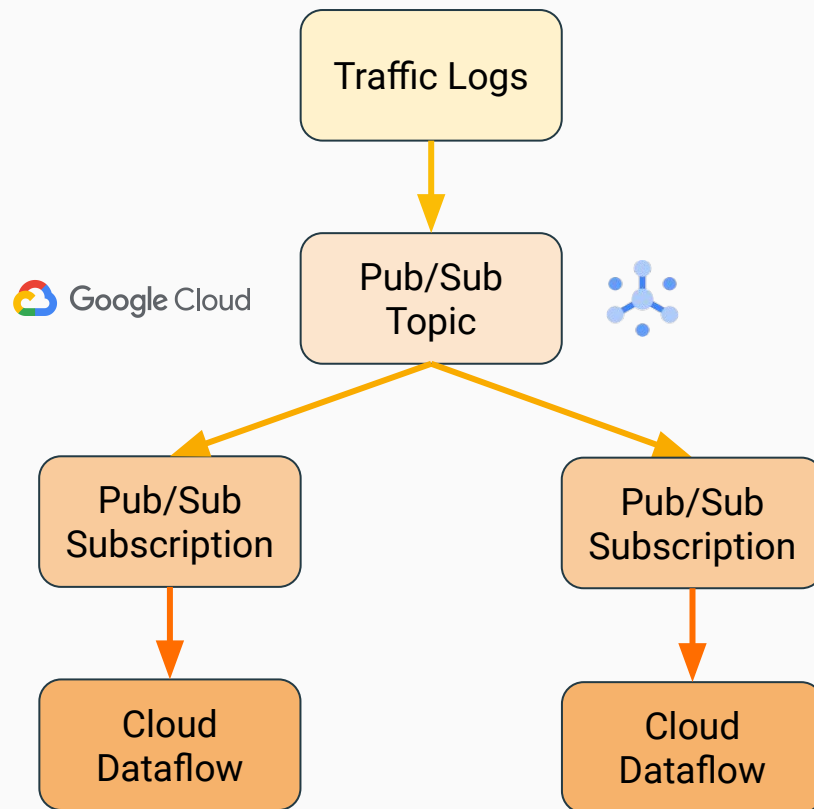


BEAM  
SUMMIT



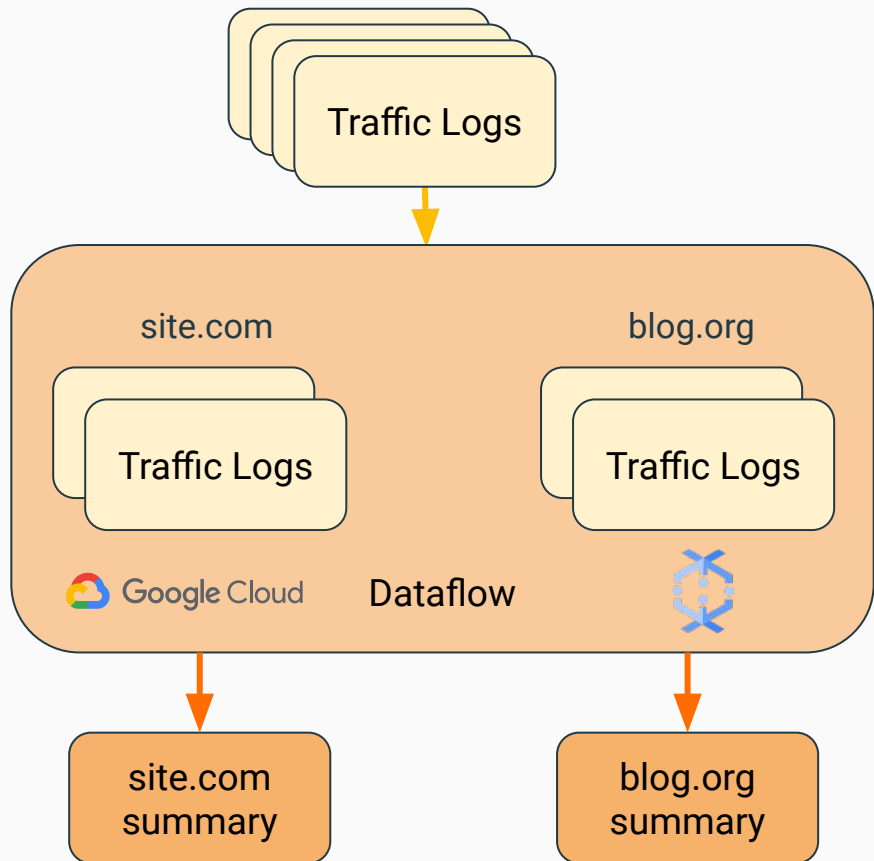
# Streaming and Pub/Sub

- Real-time analytics
- Guaranteed delivery
- Throughput metrics
- Multiple pipelines



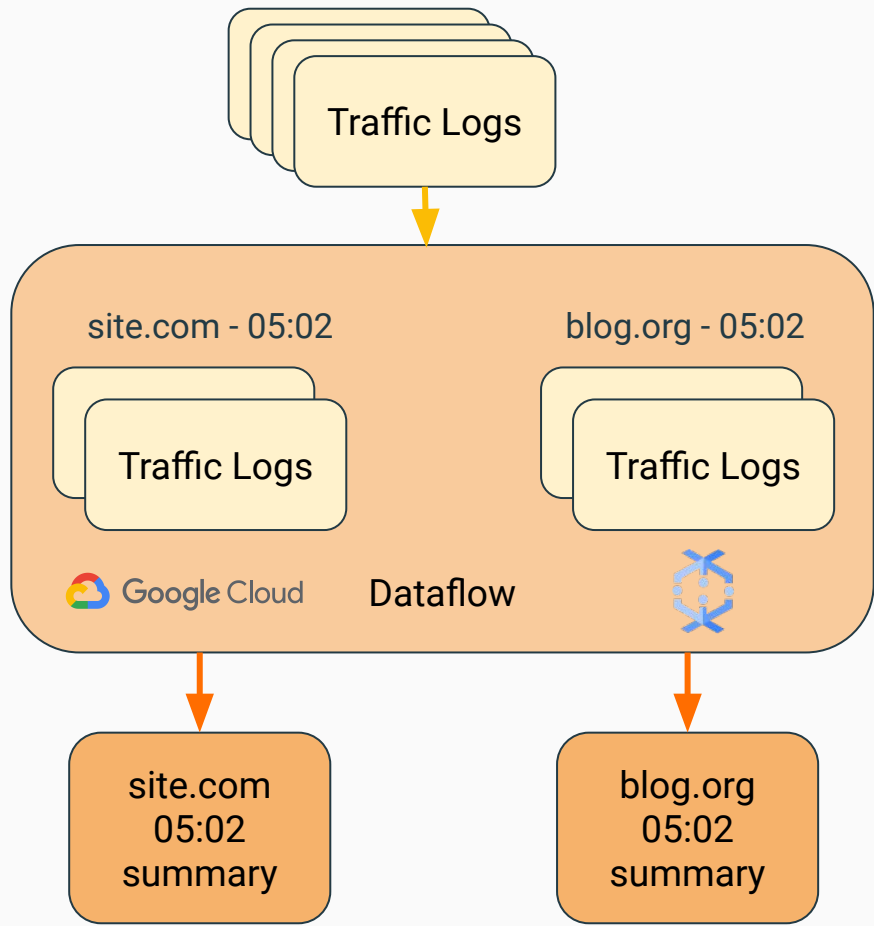
# Grouping and Reducing

- Many logs -> some reports
- Reduce to constant size
- Group by important distinguishers
- Add a key
  - beam.Map



# Windowing

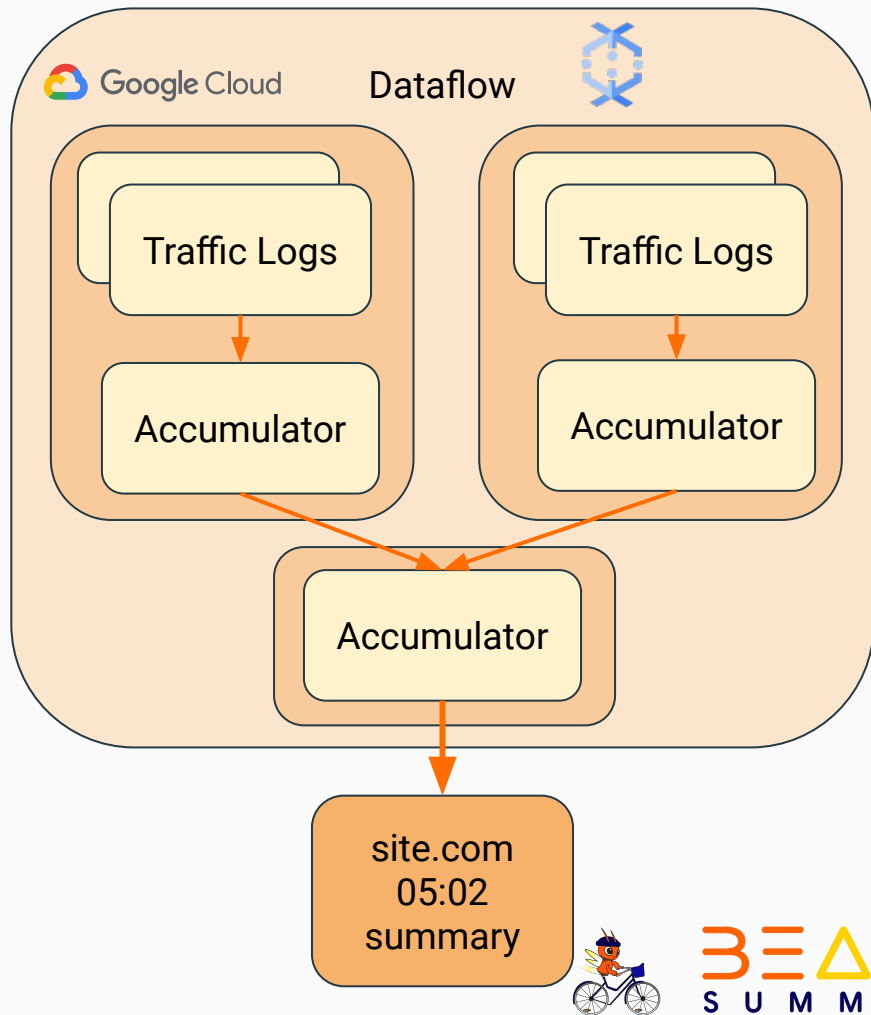
- Keep data moving
- Window by time
  - `beam.WindowInto`





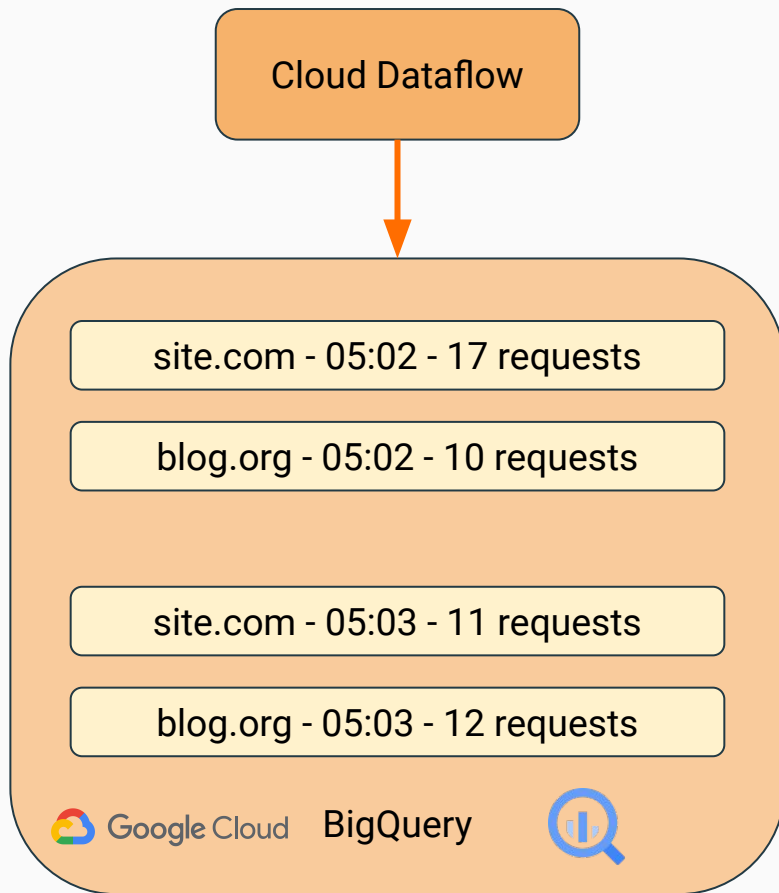
# Accumulating Results

- Process groups of data
  - `beam.CombinePerKey`
- Operate at scale
  - Summarize data
  - Merge summaries
  - Output result



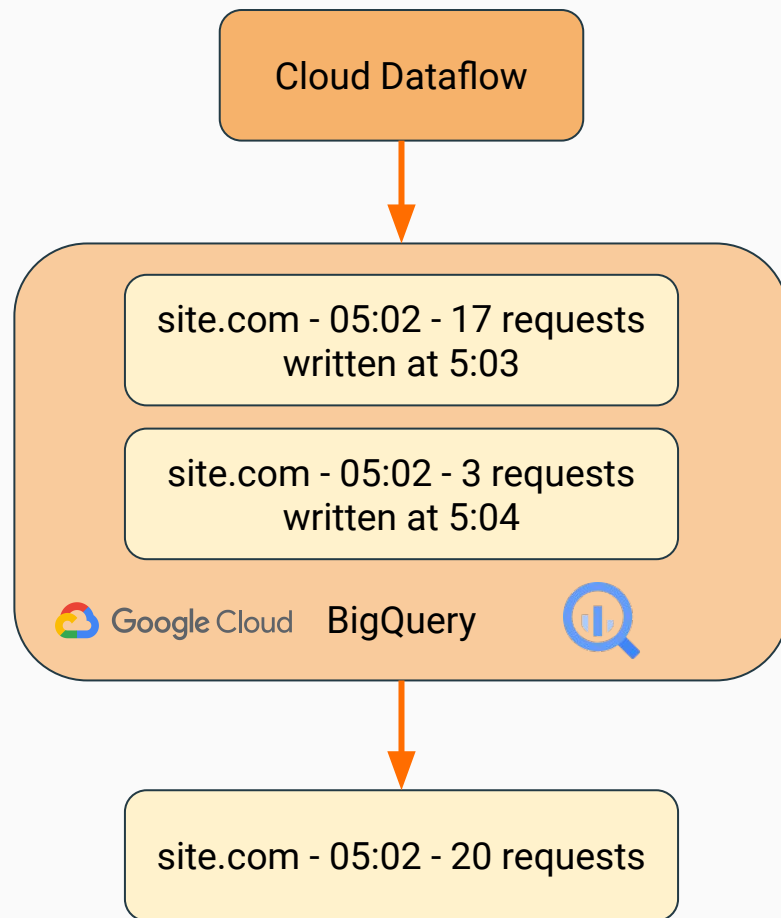
# Storing Results

- Put results somewhere
- One result per group, per window
- Many output types
- Fit needs of the service



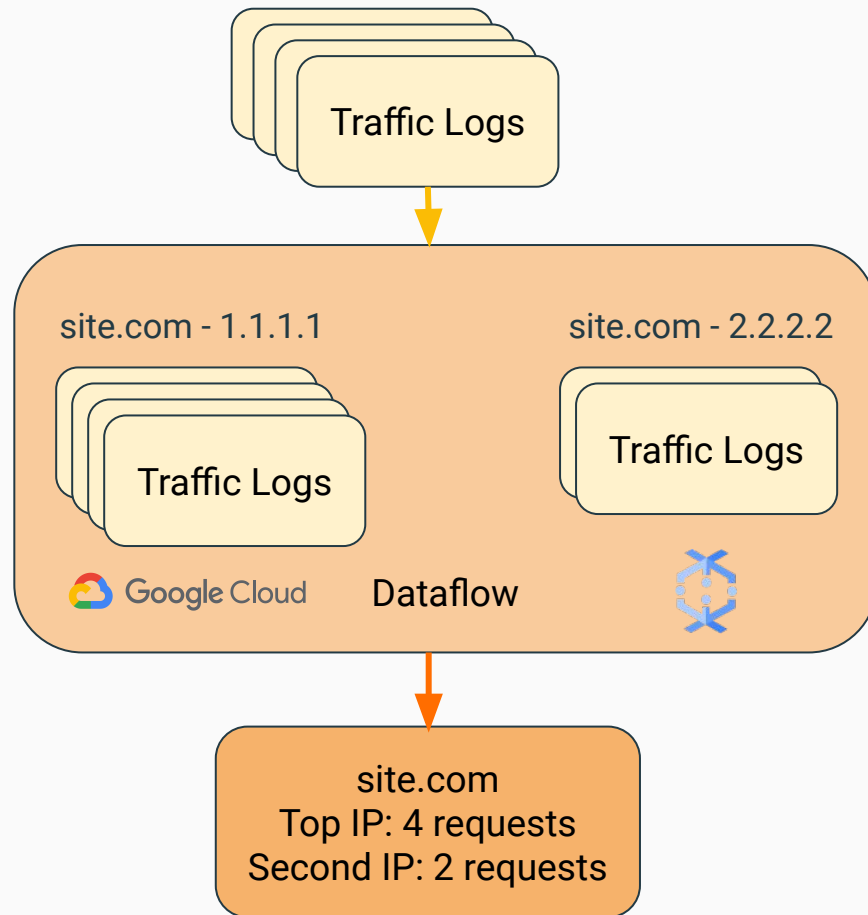
# Dealing with Stragglers

- Write their own report  
(next window) OR
- Wait for input watermark
  - `beam.transforms.window`
- Decide if data can  
combine later



# Handling PII

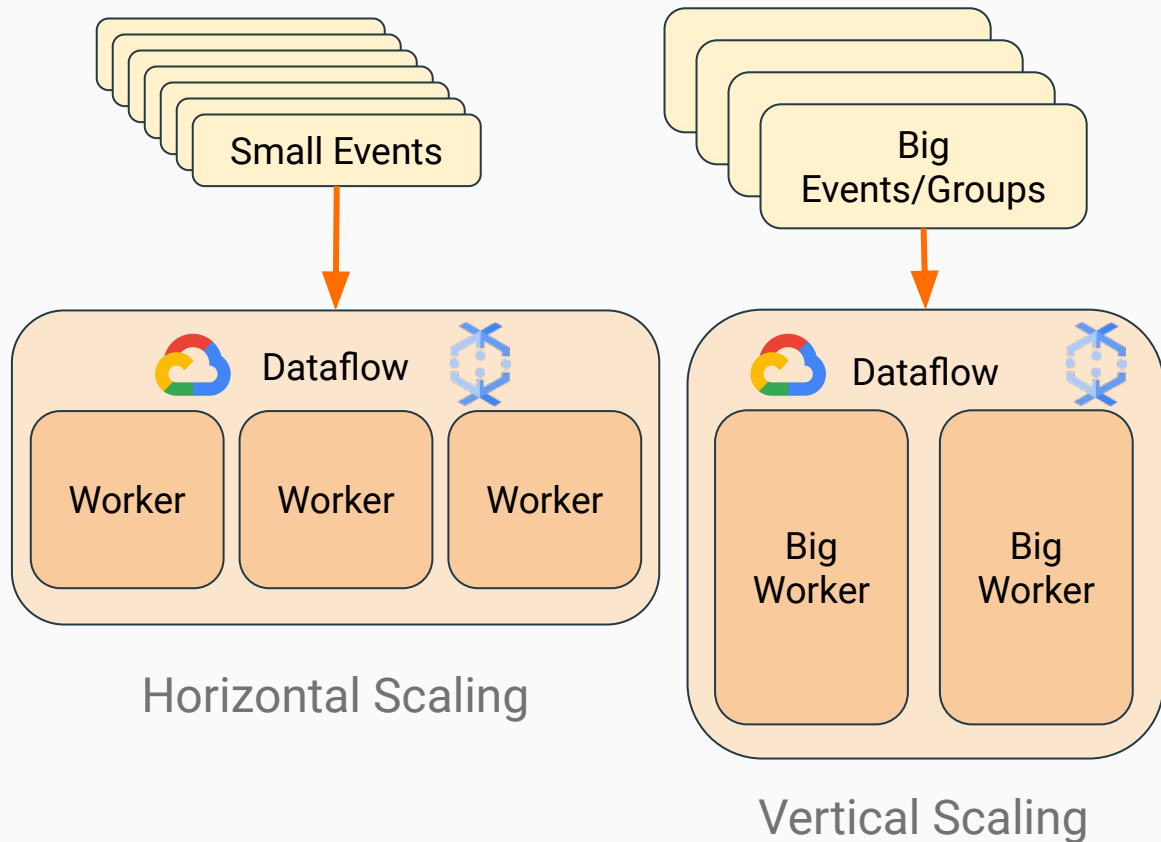
- Aggregate but don't store
- Hash top values
- Wait for watermark





# Adapting

- Scaling
  - Worker count
  - CPU / Memory per machine
- Cloud Dataflow
  - Autoscaling



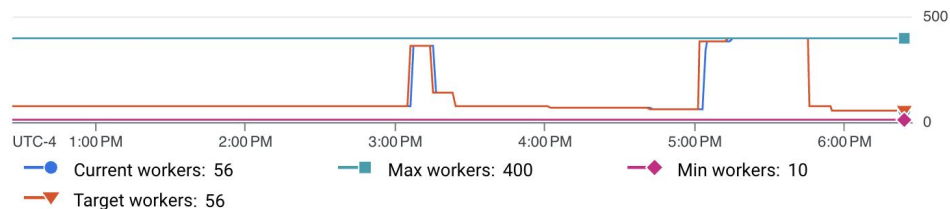
10Kbps -> 400Mbps



# Forcing Smoothness

- Baseline provisioning
- Immediate capacity
- Balance with cost

## Autoscaling ?

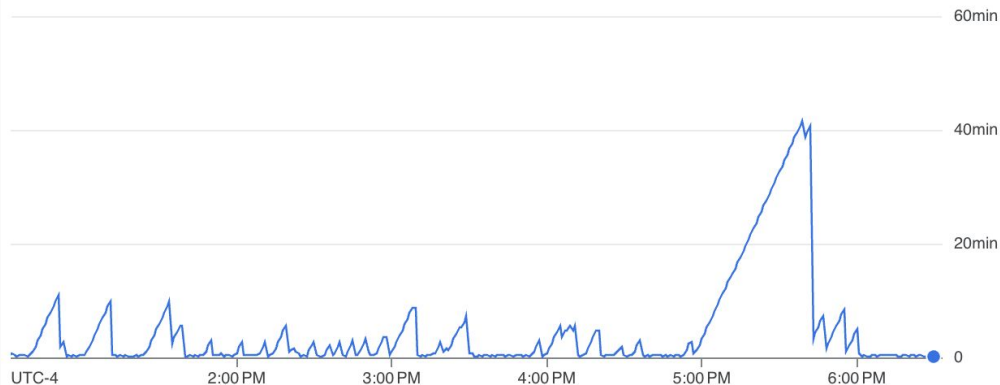


Latest worker status:

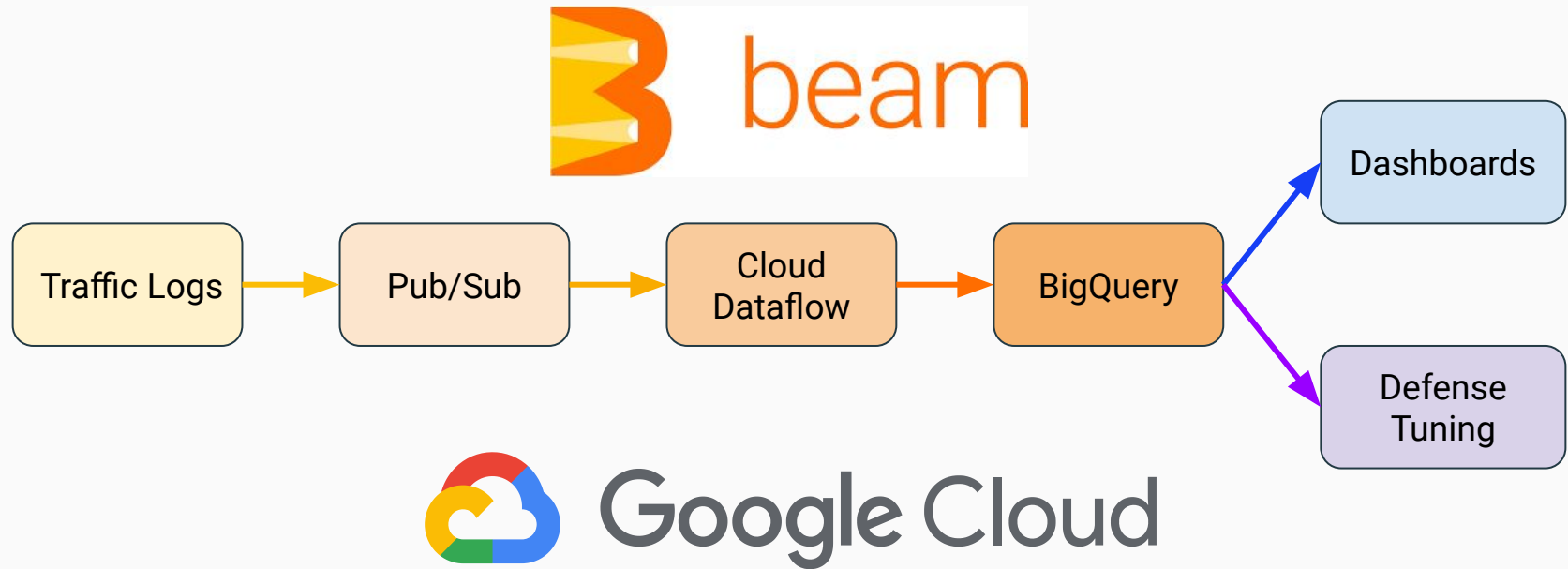
Autoscaling: Reduced the number of workers to 56 based on low average worker CPU utilization, and the pipeline having sufficiently low backlog and keeping up with input rate. The downscale magnitude was dampened to stabilize autoscaling decisions.

## Data freshness ? i

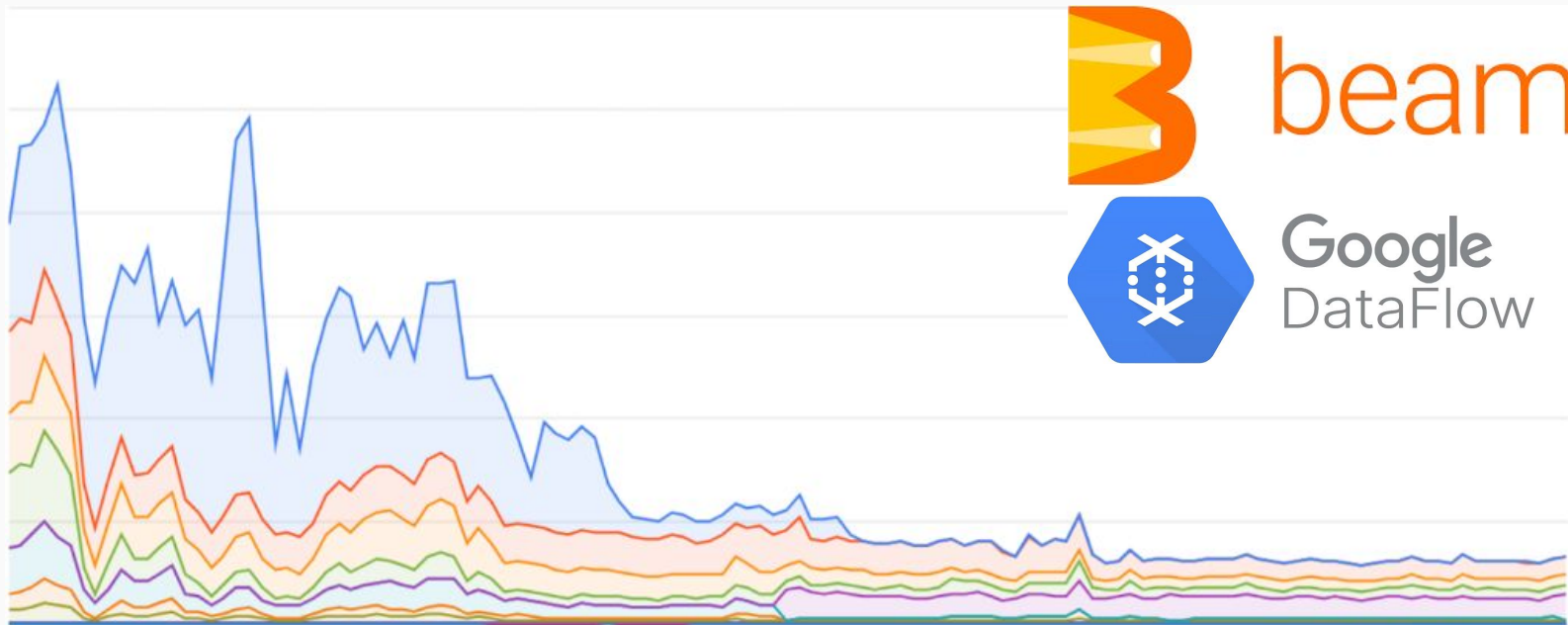
[Create alerting policy](#)



# End Result



# Reducing Costs



>90% savings over bespoke pre-processing!



BEAM  
SUMMIT



# Keeping Content Online

- Support elections
- Protect news
- Empower human rights



## Project Shield



Google Cloud



BEAM  
SUMMIT

# Thank you!

Questions?

Marc Howard

[marchoward@google.com](mailto:marchoward@google.com)



**BEAM**  
SUMMIT