

Big Data Analytics with Google Cloud Platform

Speaker



William Vambenepe
Google Cloud Platform
Lead Product Manager for Big Data

Twitter: *@vambenepe*

Google ran into
“Big Data” problems
in the process of
building its business...

Big data at Google scale



How much video is uploaded
to YouTube every **minute**?

100 hours



How many **active** Gmail
users are there?

500M+



How large is Google's web
search index?

100PB+

(over 100,000 TBs)

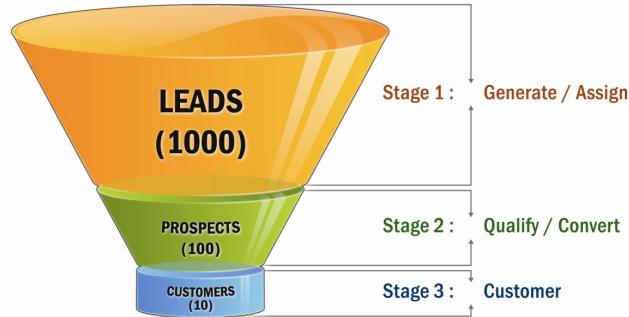


How long does it take for
Google to respond to a
search **query**?

0.25 seconds

...now that the technology
is available, its usage is
spreading to all industries.

Conventional Methods for Data & Analytics



Focused on optimizing core business processes and capture details about high interactions/transactions

Conventional Methods for Data & Analytics

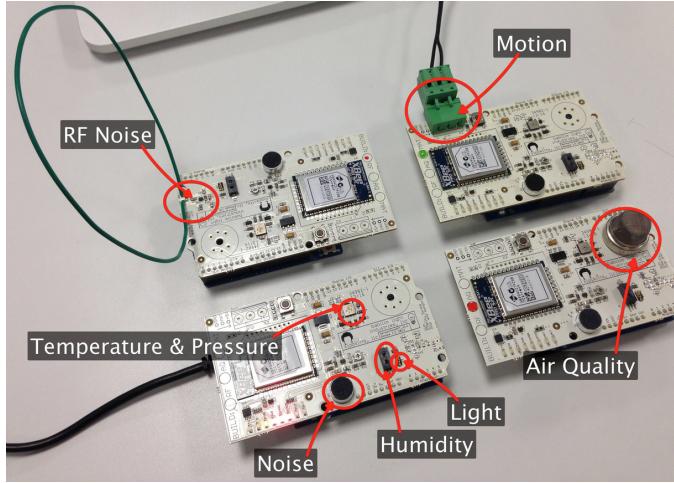
Stopped collecting data at the nearest boundary of business process:

- Not instrumented for tracking of product usage by customer
- Instrumentation for tracking supplier performance based on supplier specific situation



Businesses decision makers fill data gaps with gut feel...

Rise of Big Data enabled by Tech Innovation



It is Possible to instrument data capture at any level of business process...



...and to store and
process it easily
and efficiently.

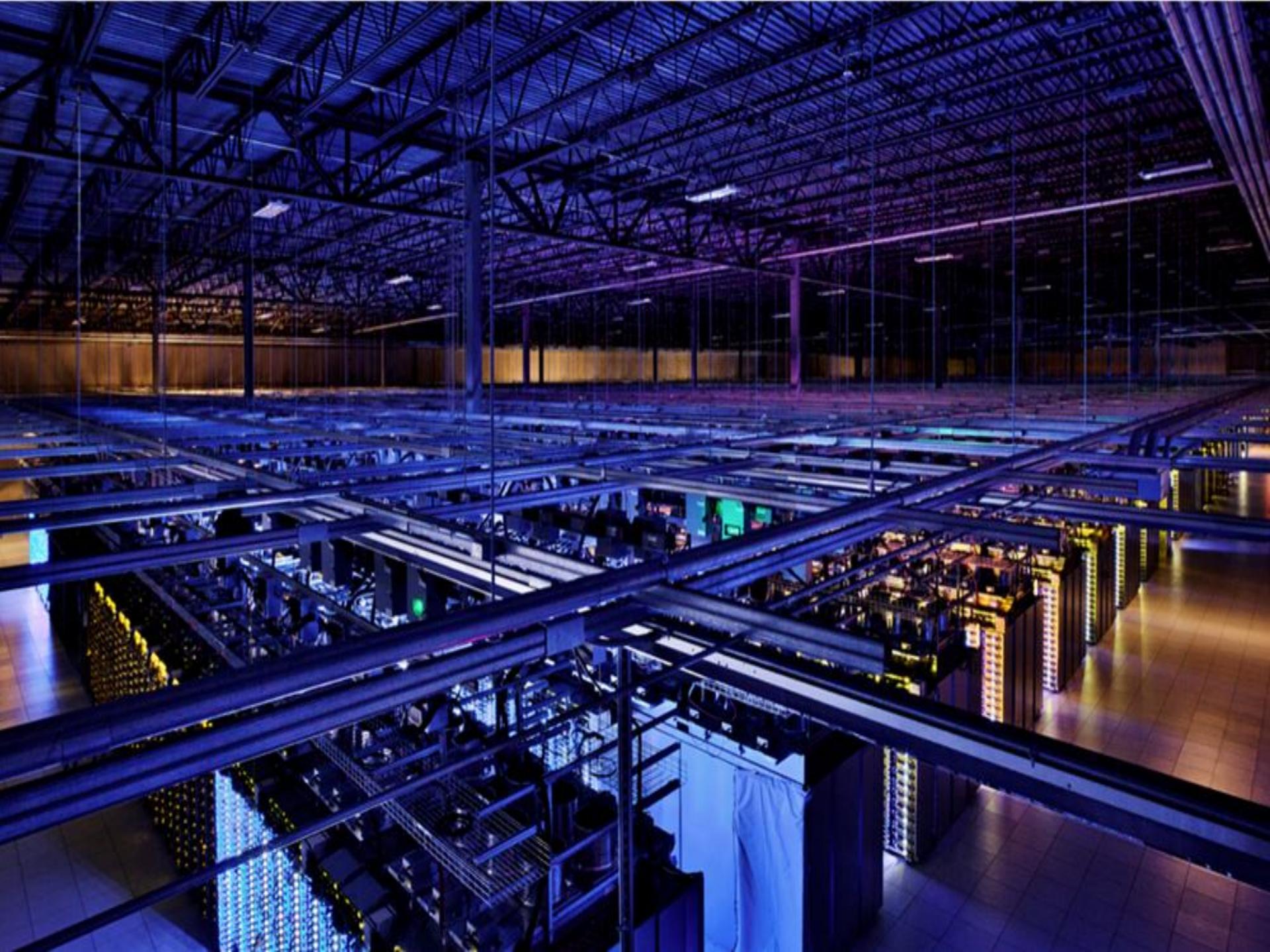
Big Data enabled by Tech Innovation



Highly scalable, performant and cost effective computational capabilities in the **cloud** makes it possible to store vast amounts of data from many sources



For the past **15 years**, Google has been building out the world's fastest, most powerful, highest quality cloud infrastructure **on the planet**.



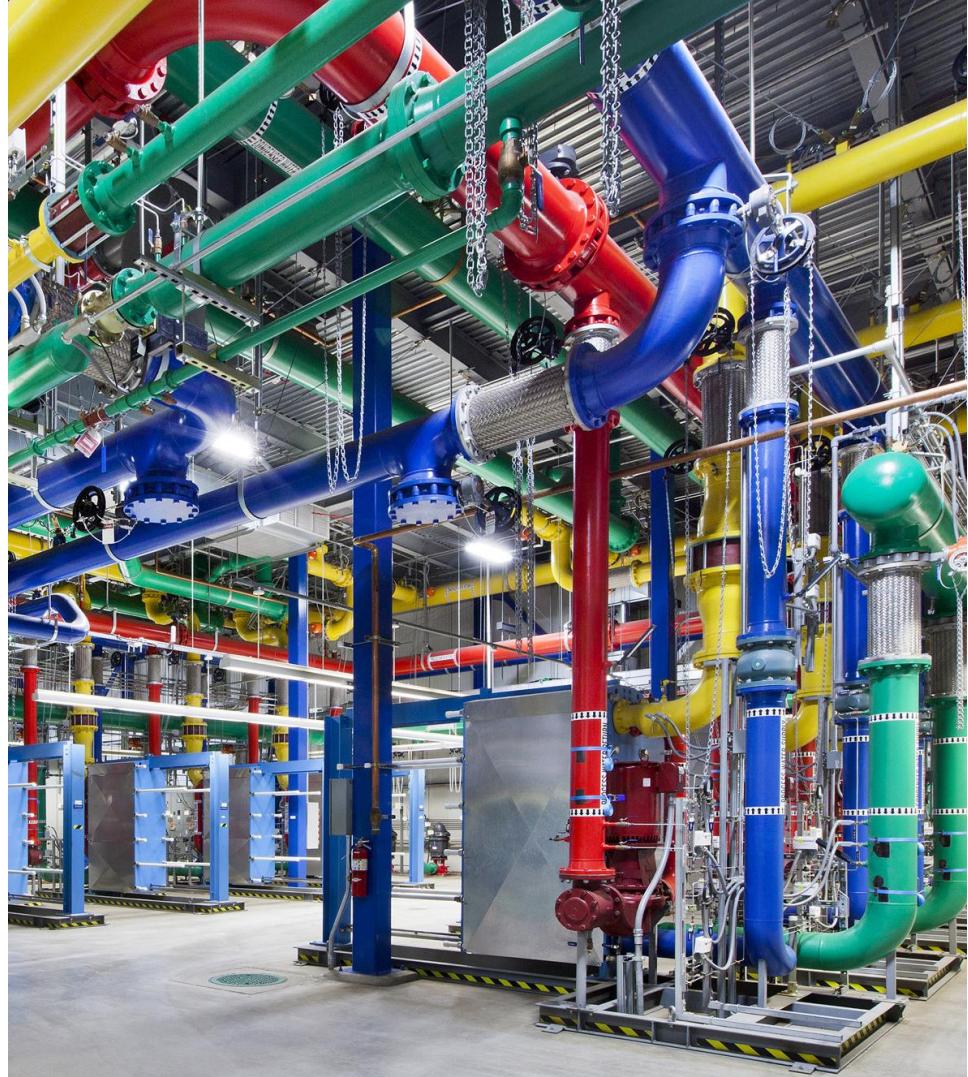
Google's Platform

"[Google's] ability to build, organize, and operate a huge network of servers and fiber-optic cables with an efficiency and speed that rocks physics on its heels.

This is what makes Google Google: its physical network, its thousands of fiber miles, and those many thousands of servers that, in aggregate, add up to the **mother of all clouds.**"

- *Wired*

Wired, 'Google Throws Open Doors To Its Top Secret Data Center', October 2012



Images by Connie Zhou

Investing In Our Cloud

\$2.9B in additional data center investments worldwide



Googling good economic news for our community and state

Bloomberg

Google to Build Taiwan Center as Demand for Gmail, YouTube Rises

Chicago Tribune

Google to invest 150 mln euros in Finland data centre



Google to invest nearly \$400 million in Belgian data center to 'meet growing demand for its online services'

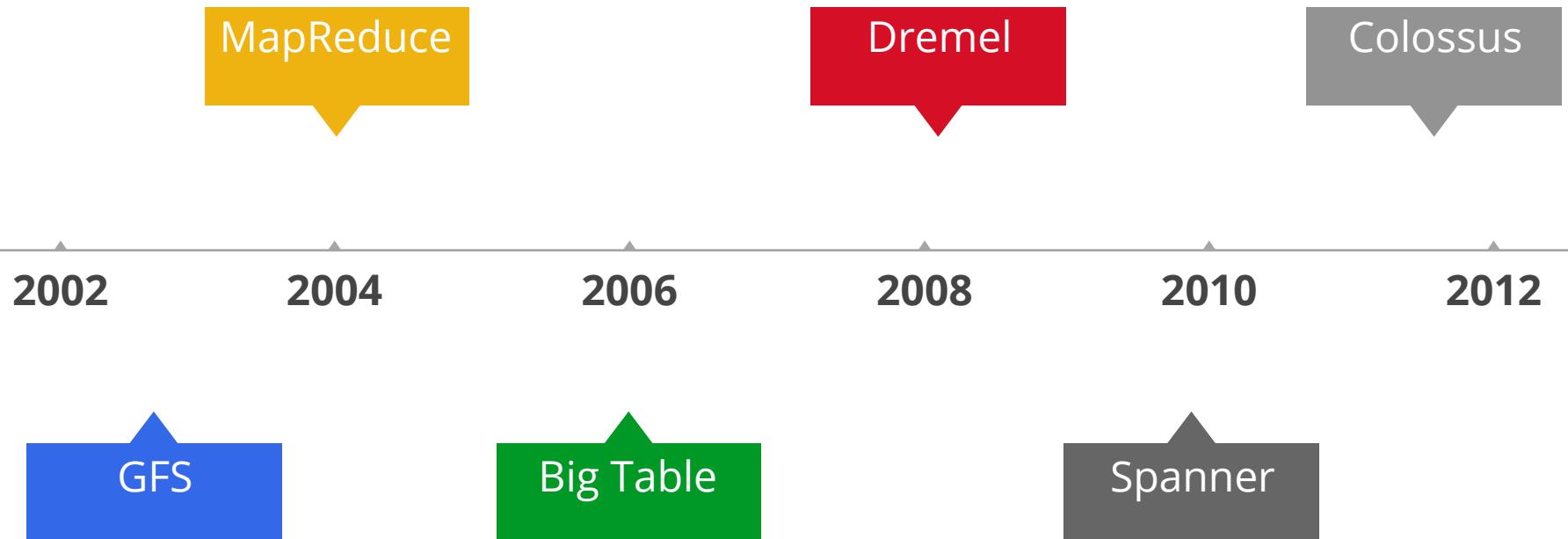


Google Expands into Latin America With New Data Center in Chile

THE WALL STREET JOURNAL.

Google to Invest \$120 Million in Singapore Data Center

Google Innovations in Software



Google Cloud Platform

Compute



Compute
Engine



App Engine

Storage



Cloud Storage



Cloud SQL



Cloud
Datastore



Persistent
Disk

App Services



BigQuery



Cloud
Endpoints



Caching



Queues

Google Compute Engine and Google Cloud Storage



Google Compute Engine



Compute Engine

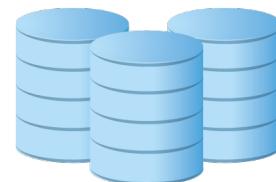
Virtual Machine Hosting



Compute



Network



Storage



Tooling

Launch Virtual
Machines on demand

Connect your VMs
together to form
powerful clusters

Store on persistent
disk, local disk or
Cloud Storage

Control your VMs via
REST API or
command line

Scale

Performance

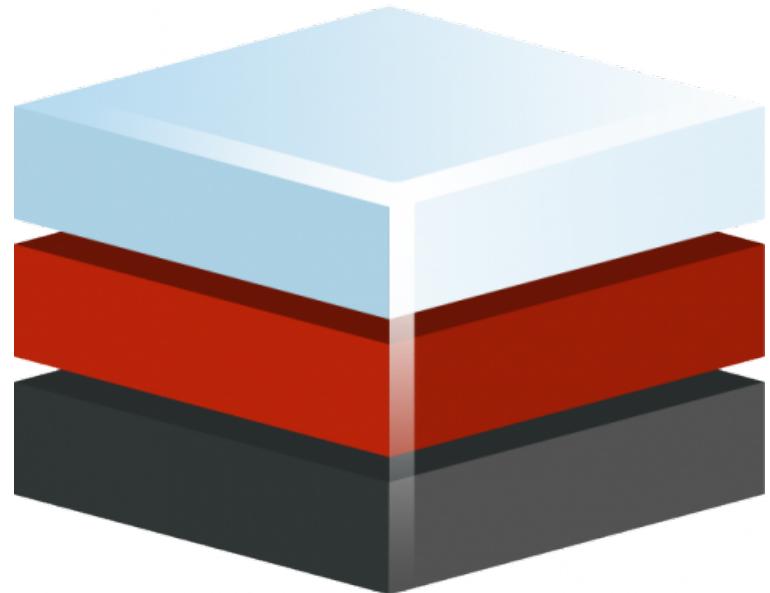
Value

Google Compute Engine

“

“Google Compute Engine is not just fast. It’s Google fast. In fact, it’s a class of fast that enables new service architectures entirely.”

- Sebastian Stadil, Scalr



Google Compute Engine

MapR on GCE breaks the TeraSort, then MinuteSort records

- sorted 15 billion 100-byte records (1.5TB) in 59 seconds
- used 2,103 instances (n1-standard-4-d)
- 8,412 cores

Computing in the Cloud with GCE



Cancer Research

"Google Compute Engine is just part of what we see as a whole new way for scientists around the world to **work more effectively in the cloud**. This is a paradigm shift that Google services can help bring about."

— *Hector Rovira, senior software engineer at the Institute for Systems Biology*

High Performance Computing

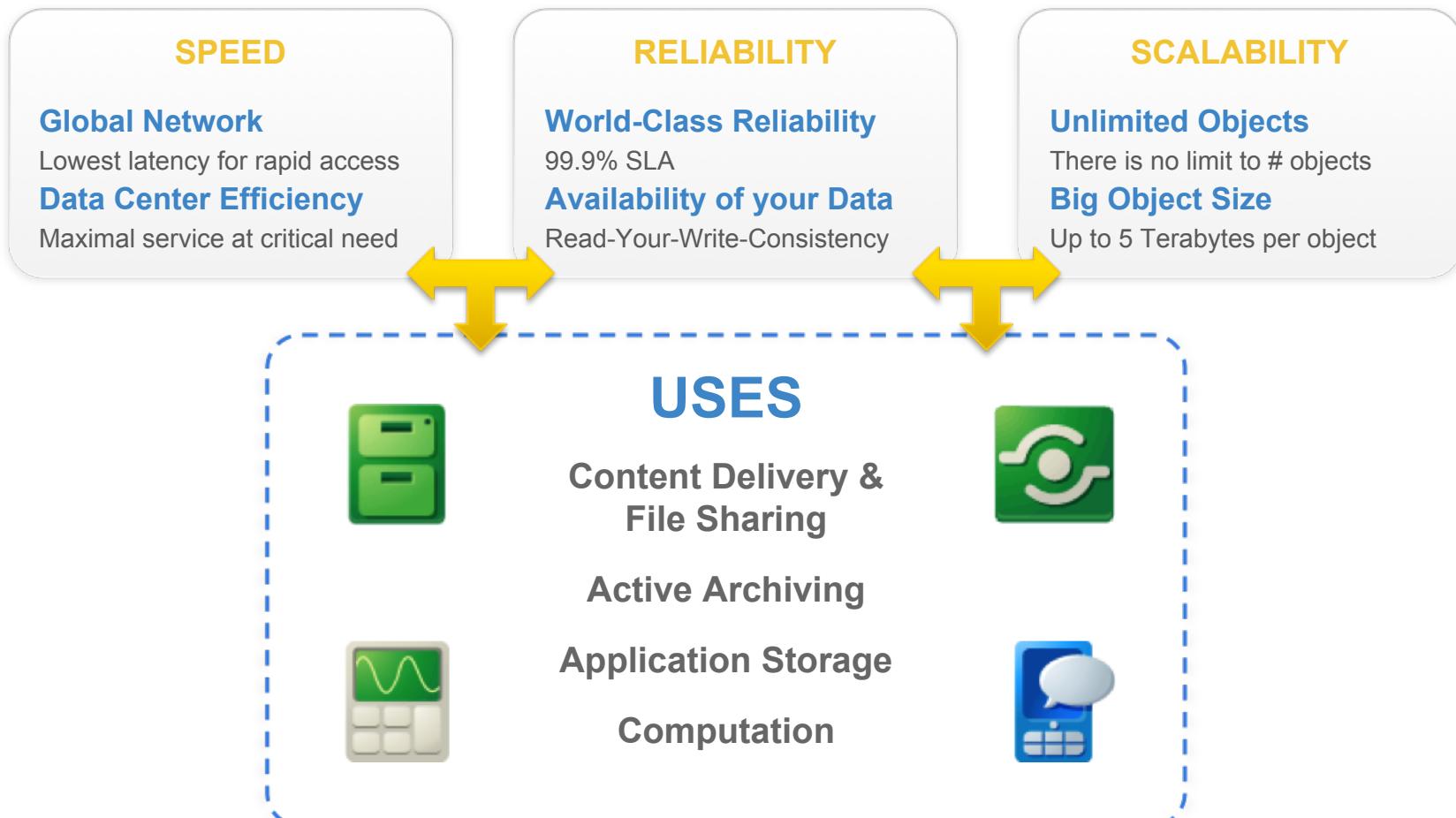
"And we have been very satisfied with the way Google Compute Engine performs. One of the great things with Google Compute Engine in our experience is that **the performance is really reliable**, so there is not a lot of variability in the performance we see."

— *Joe Masters Emison, Founder and VP R&D, BuildFax*

Google Confidential, Do not Copy or Distribute

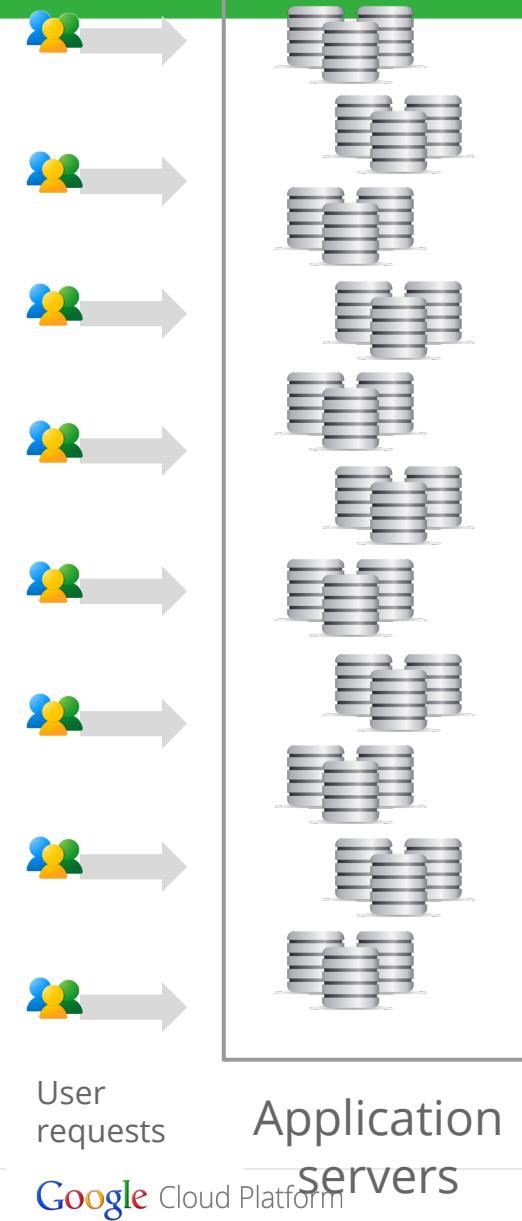
Google Cloud Storage

Google Cloud Storage lets you store and access data on Google's reliable infrastructure

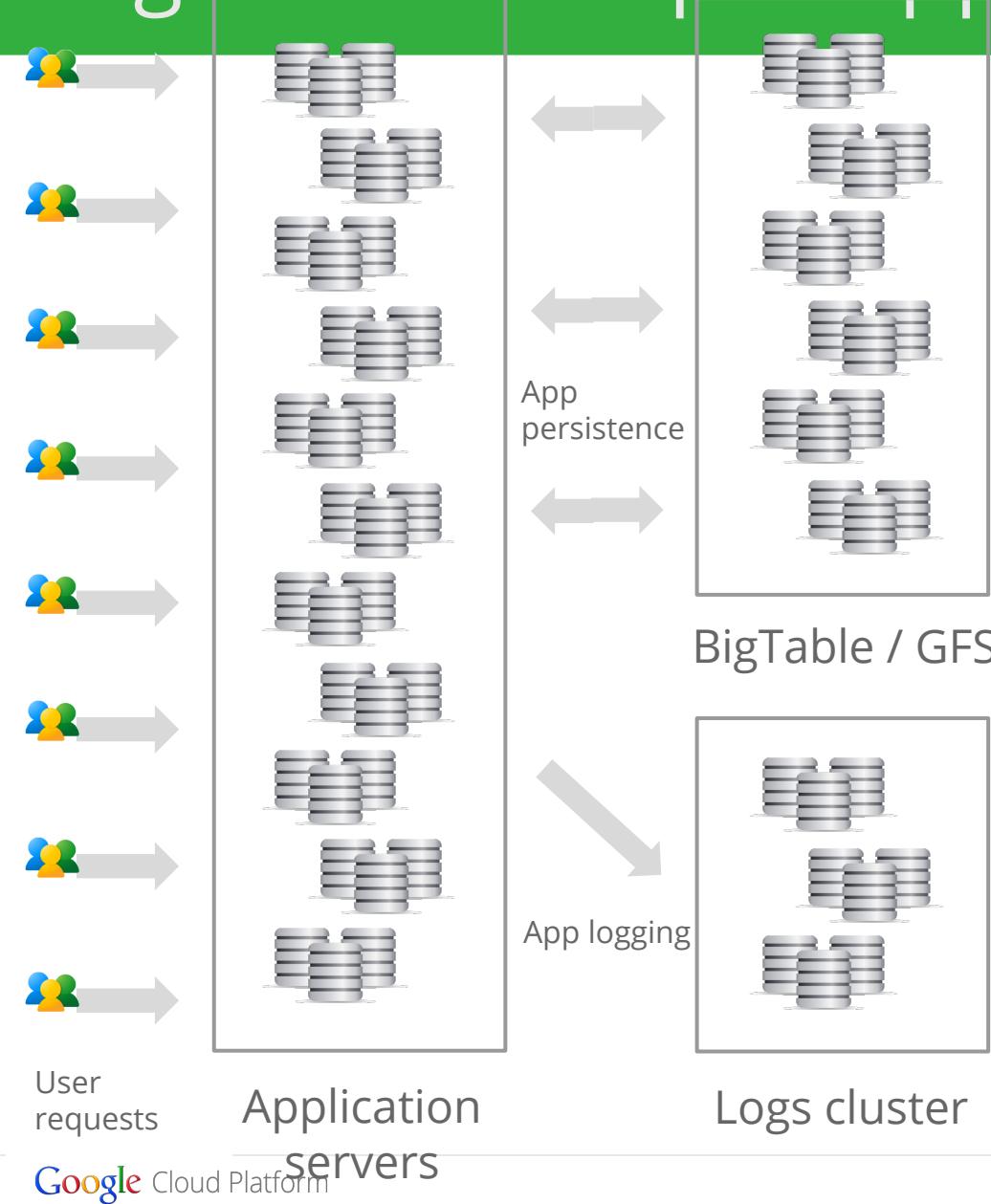


Specialized Big Data services of **Google** Cloud Platform

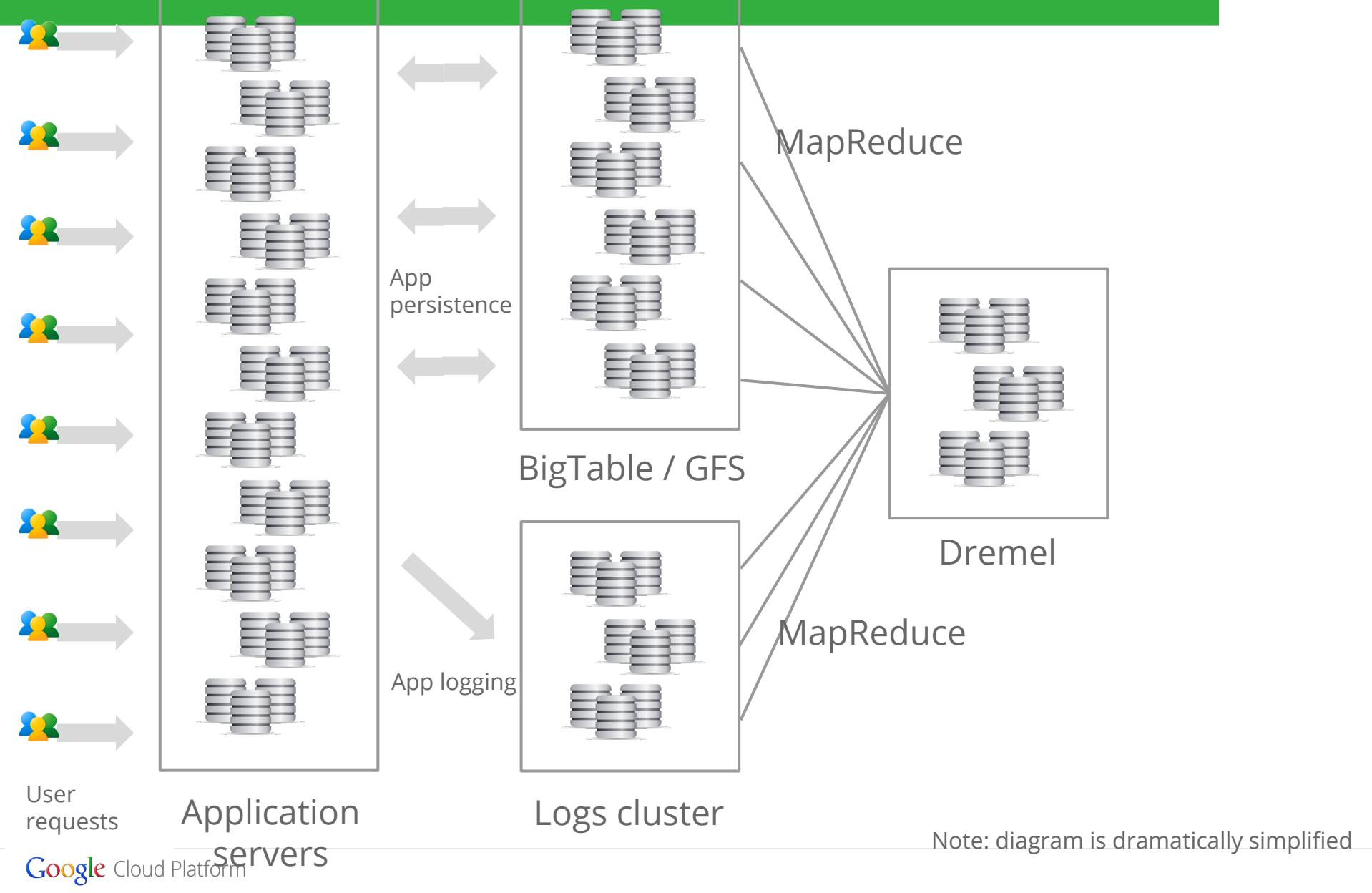
Big Data example: app logs



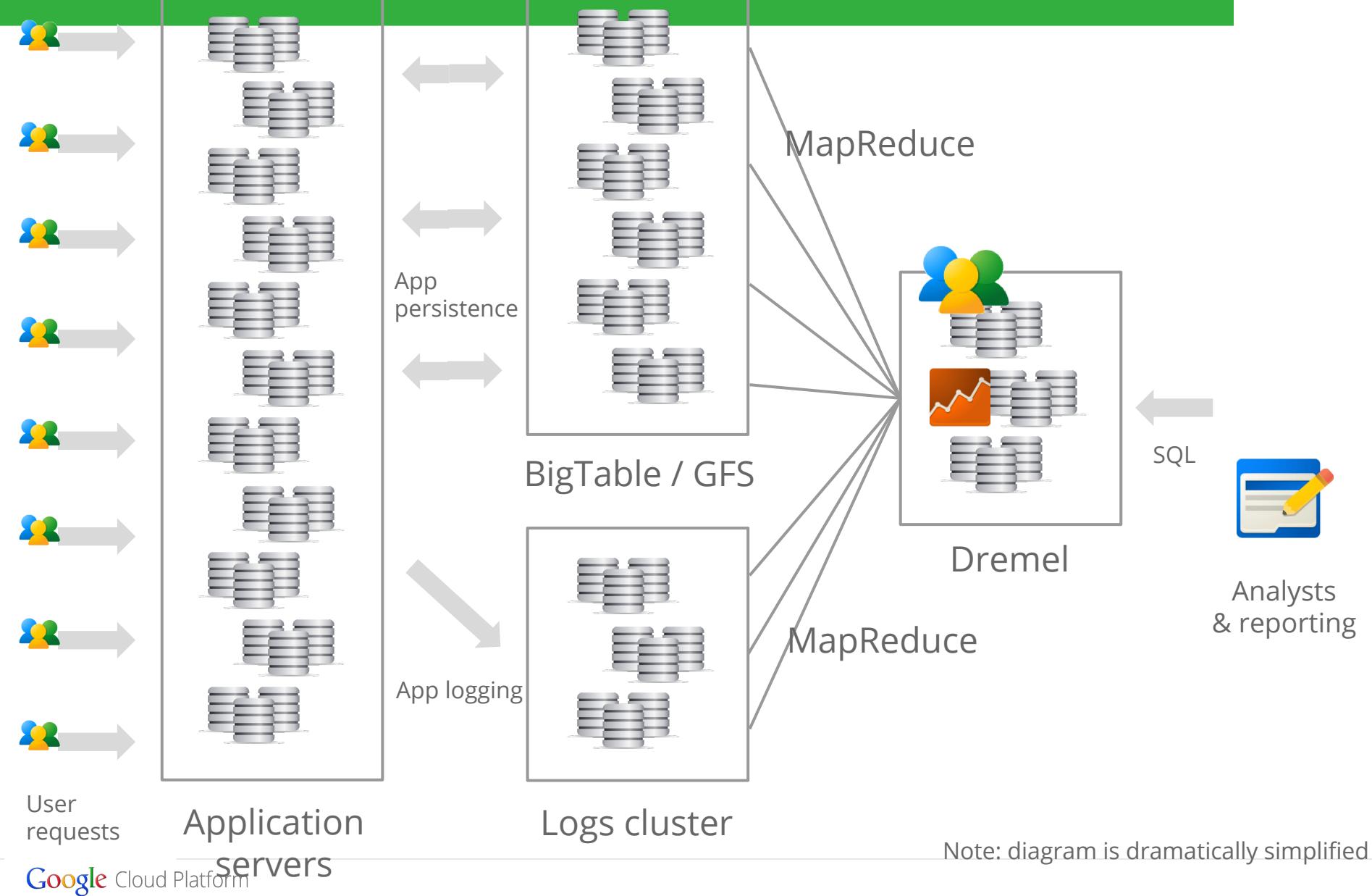
Big Data example: app logs



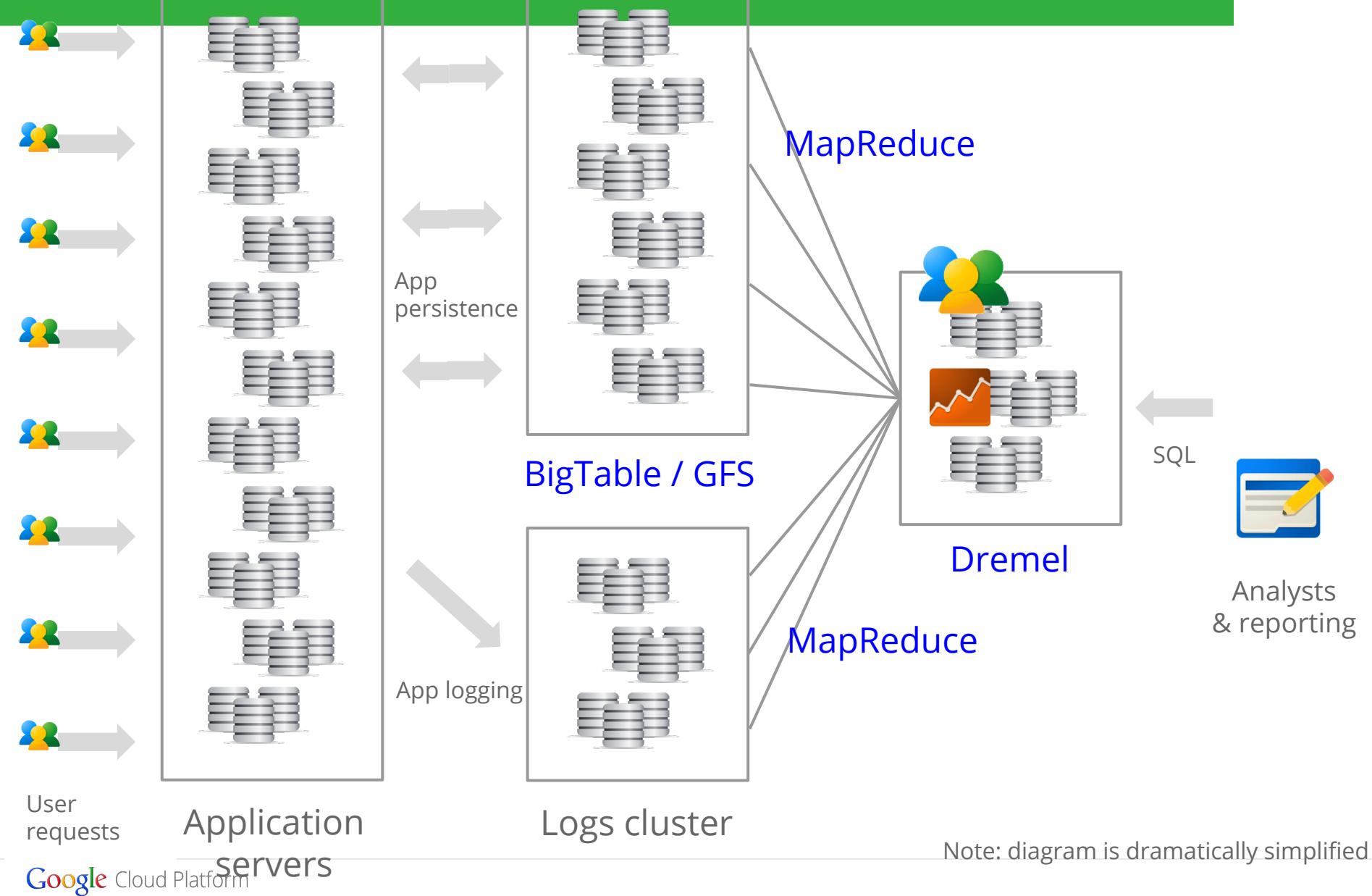
Big Data example: app logs



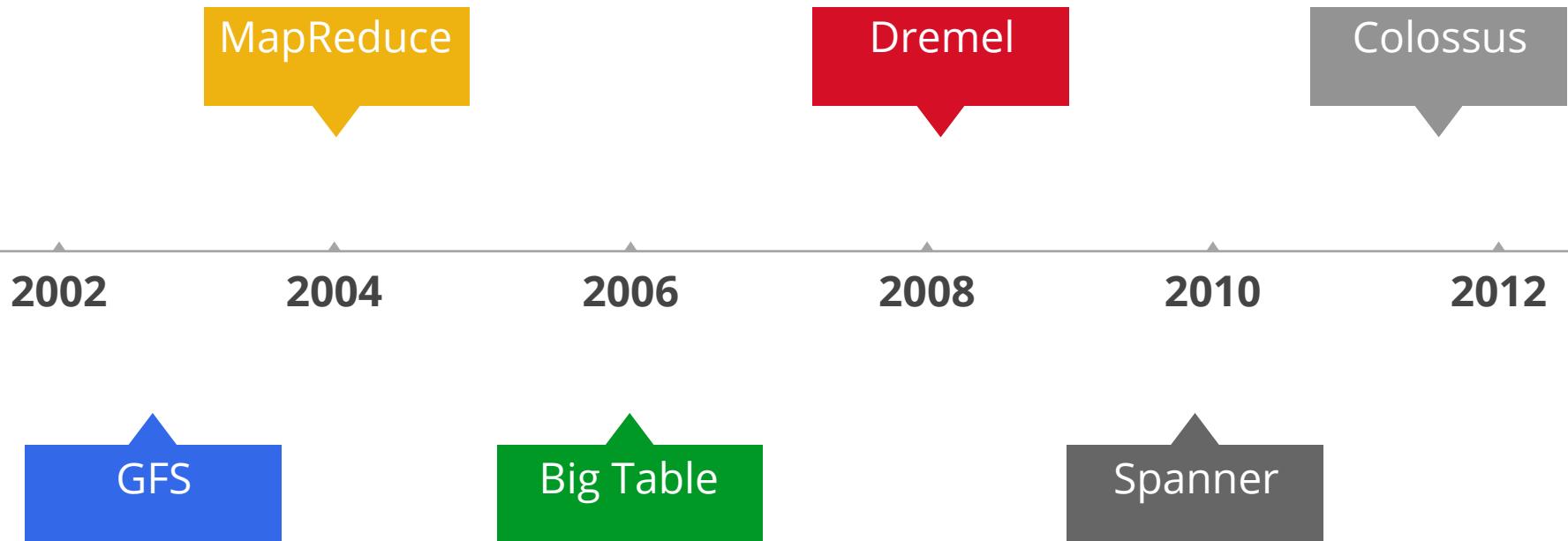
Big Data example: app logs



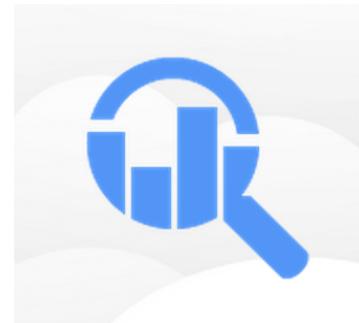
Big Data example: app logs



Google Innovations in Software



Google Cloud Platform Services: BigQuery



Google BigQuery

- Query billions of rows in seconds, with no index
- Uses a SQL-style query syntax
- Bulk and stream data ingestion
- As a service:
 - zero setup/management
 - accessed by a
 - RESTful API
 - console
 - partner tools
 - Pay as you go (per query)

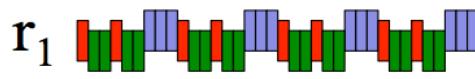


How Google Approaches Analytics

Dremel

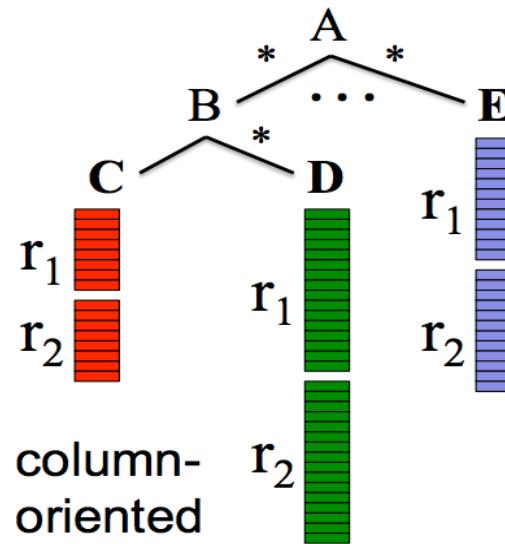
Ad-hoc query system for terabyte datasets

- Query execution tree
- Column Oriented records
- ...and a lot of nodes



...

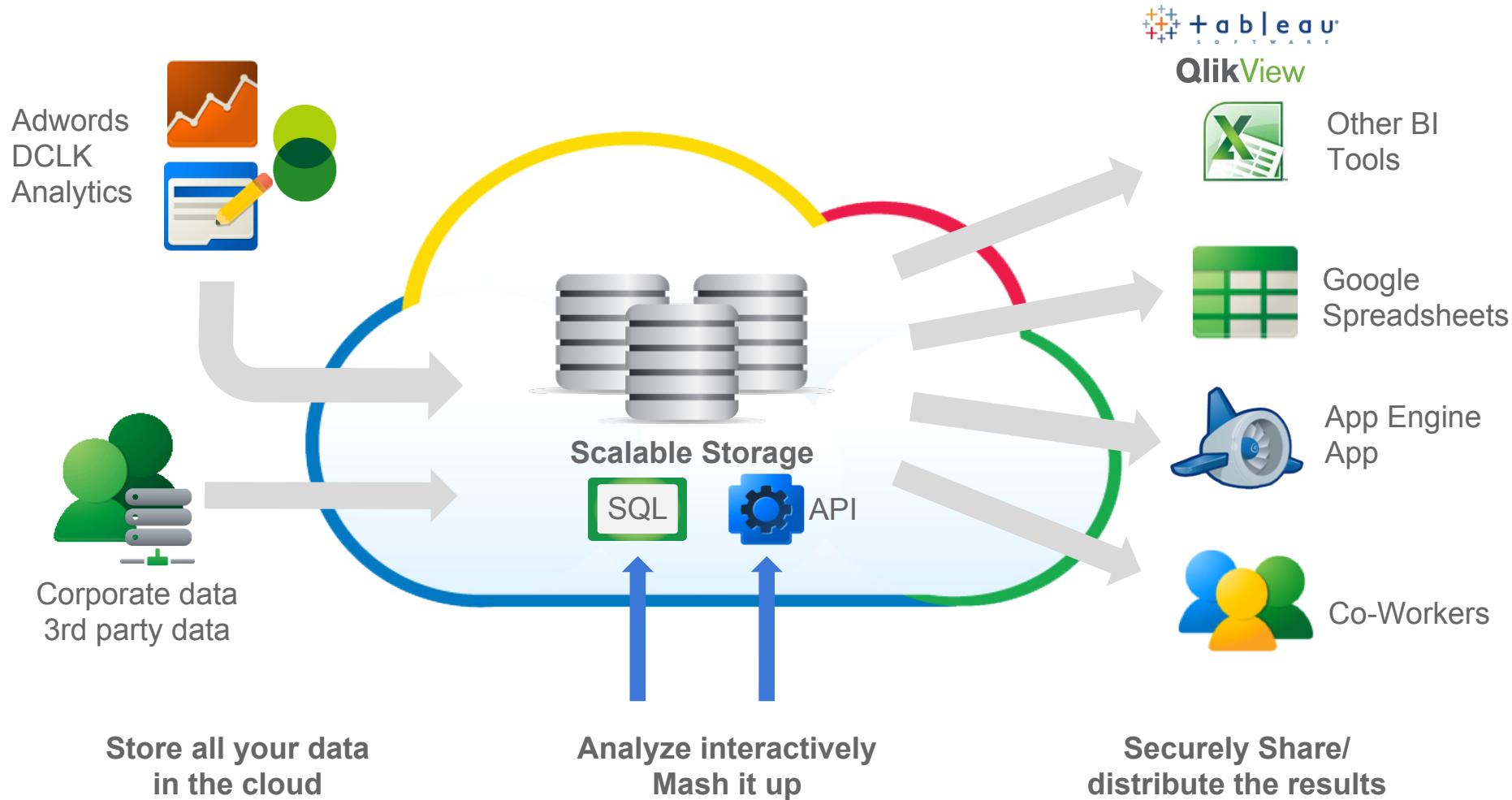
record-
oriented



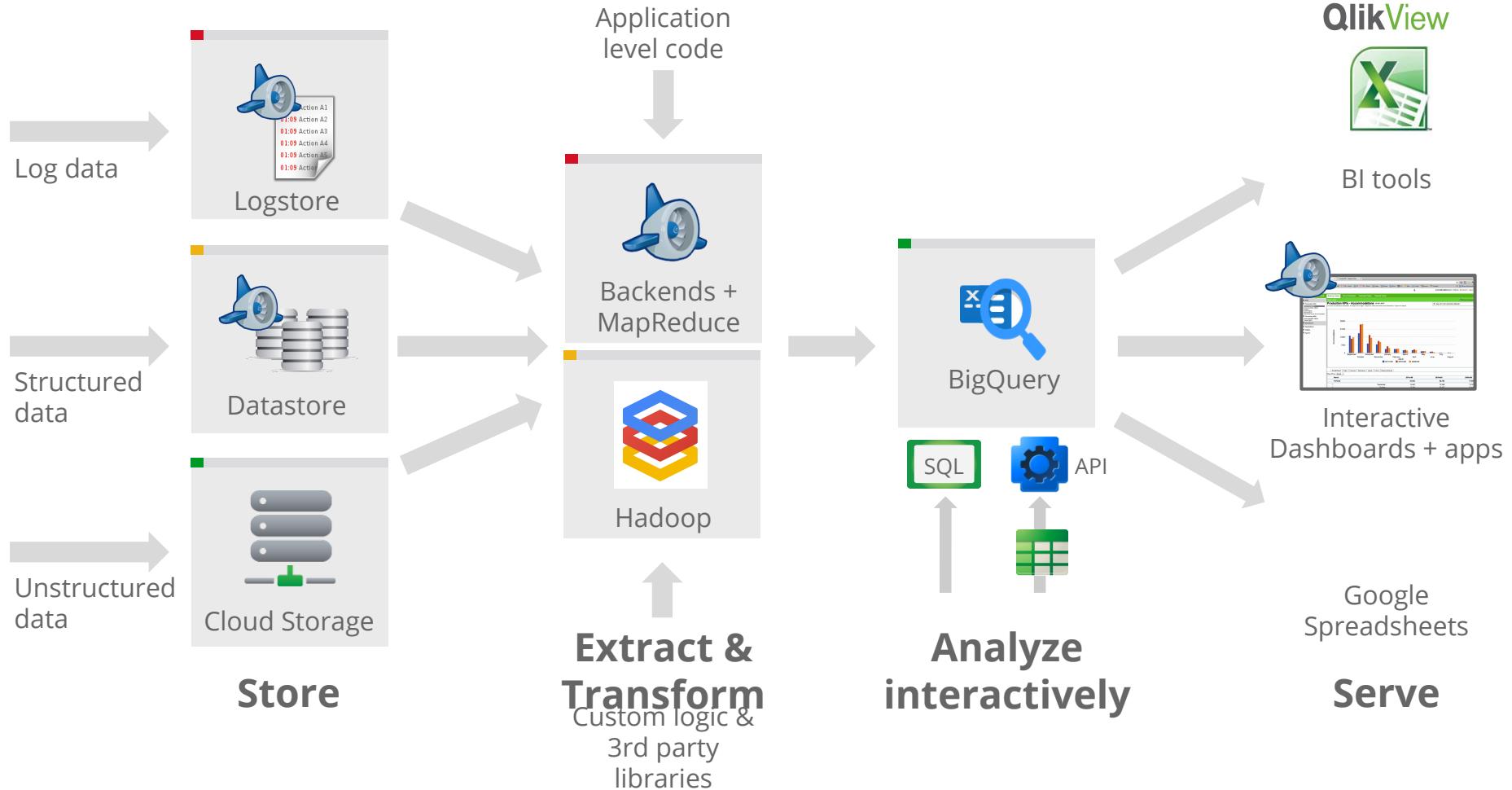
Google BigQuery

BigQuery: a fully-managed data analytics service in the cloud.

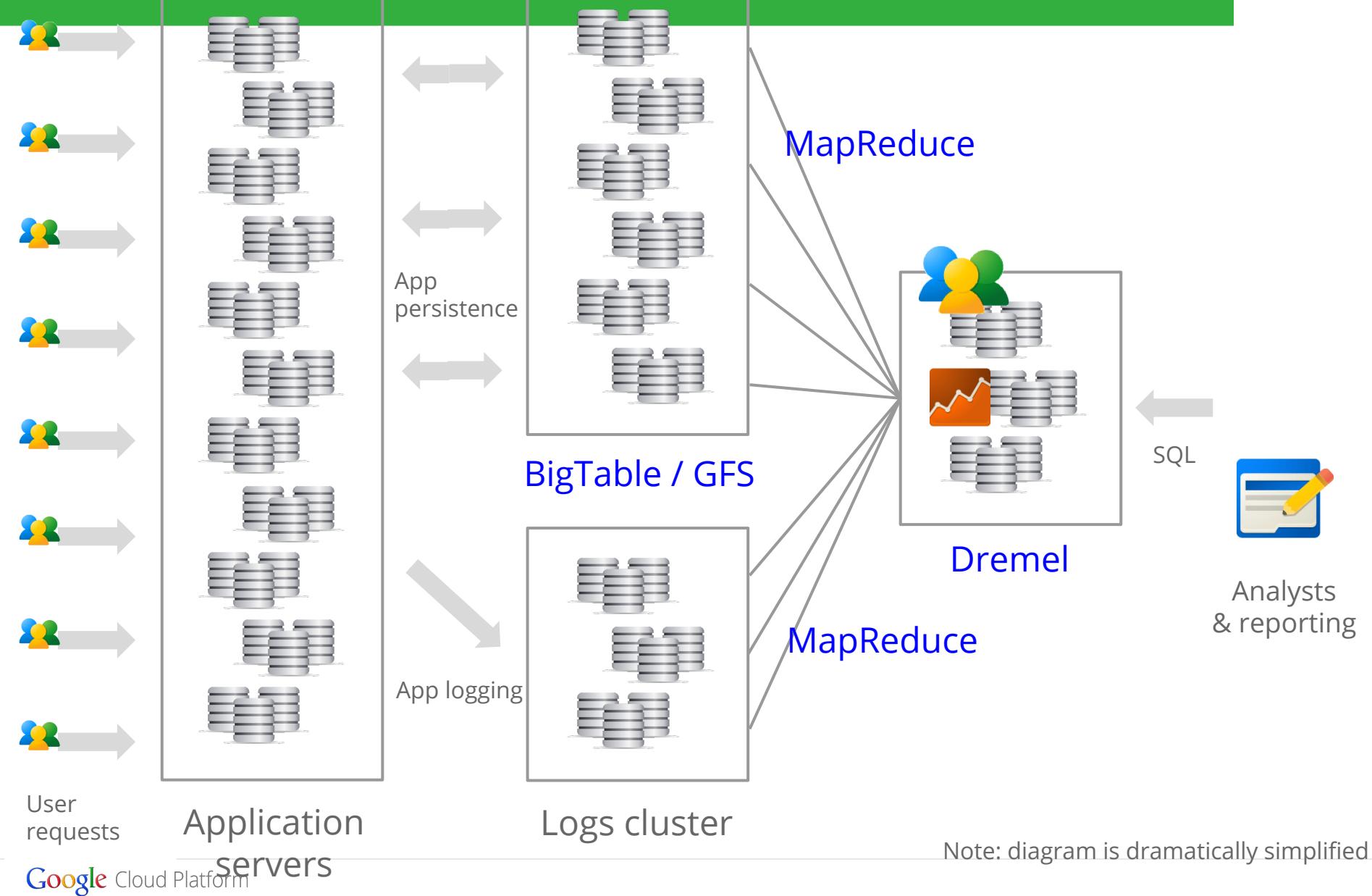
Unlimited storage. Interactive analysis on multi-terabyte datasets.



Big Data Processing Pipeline



Big Data example: app logs



Sample BigQuery use cases

Display ads analytics

(global top-5 media agency)

Analyze global campaign performance for F500 clients
1 client = 20GB/day of DoubleClick impressions logs

Ads network reporting

(3rd party mobile ads)

Deliver x-platform performance analytics dashboards
1B events/day x 100s of ads customers

Fleet reservations

(online travel operations)

Monitor customer demand vs supply shortfalls
10,000 routes x 1000s customers = millions of daily events

Mobile app statistics

(online reading vendor)

Usage analysis on 60M installs; 10M active users
2B API requests/day, 20GB log data/day

Revenue optimization

(holiday/travel properties)

Correlate marketing effectiveness vs global reservations
10MBs / day from multiple data warehouses

\$2000 to get started with Google Cloud Platform

Google Cloud Platform Starter Pack allows developers from affiliated partners to receive \$2,000 of credit - \$1,000 for Google App Engine and \$1,000 for Google Compute Engine.

<https://cloud.google.com/resources/starterpack/>

\$2,000 in Google Cloud Platform
credit to launch your idea



\$1,000 for App Engine



\$1,000 for Compute Engine

Claim your credit and learn more at [cloud.google.com/starterpack](https://cloud.google.com/resources/starterpack)

Promo code