

# Google Cloud - ATLAS



Presented by:

Karan Bhatia, SA Manager, [karanbhatia@google.com](mailto:karanbhatia@google.com)

Usman Qureshi, CE, [googleusman@google.com](mailto:googleusman@google.com)

Google Cloud

# What have been up to the past week...



 Sundar Pichai   
@sundarpichai

We're partnering w/ the federal govt + other leaders in technology & academia to provide @GoogleCloud high performance computing resources to researchers working on potential #COVID19 treatments & vaccines. Researchers can apply at [covid19-hpc-consortium.org](https://covid19-hpc-consortium.org)




WH.GOV

White House Announces New Partnership to Unleash U.S. Supercomputing Resources to Fi...  
[whitehouse.gov](https://whitehouse.gov)



 Sundar Pichai   
@sundarpichai

We're rolling out Teach from Home, a tools & resources hub to help teachers continue teaching during school closures. We're also establishing a \$10M Distance Learning fund to support orgs that help remove barriers for students learning remotely.  
[#COVID19](https://www.google.com/teachfromhome)




Teach from Home



Helping educators and students stay connected  
[blog.google](https://blog.google)

 Sundar Pichai   
@sundarpichai


The @DeepMind team is sharing predictions of some protein structures associated with the virus that causes COVID-19, generated by the latest AlphaFold system. We hope this will help researchers around the world in their current efforts.



Computational predictions of protein structures associated with COVID-19  
[deepmind.com](https://deepmind.com)

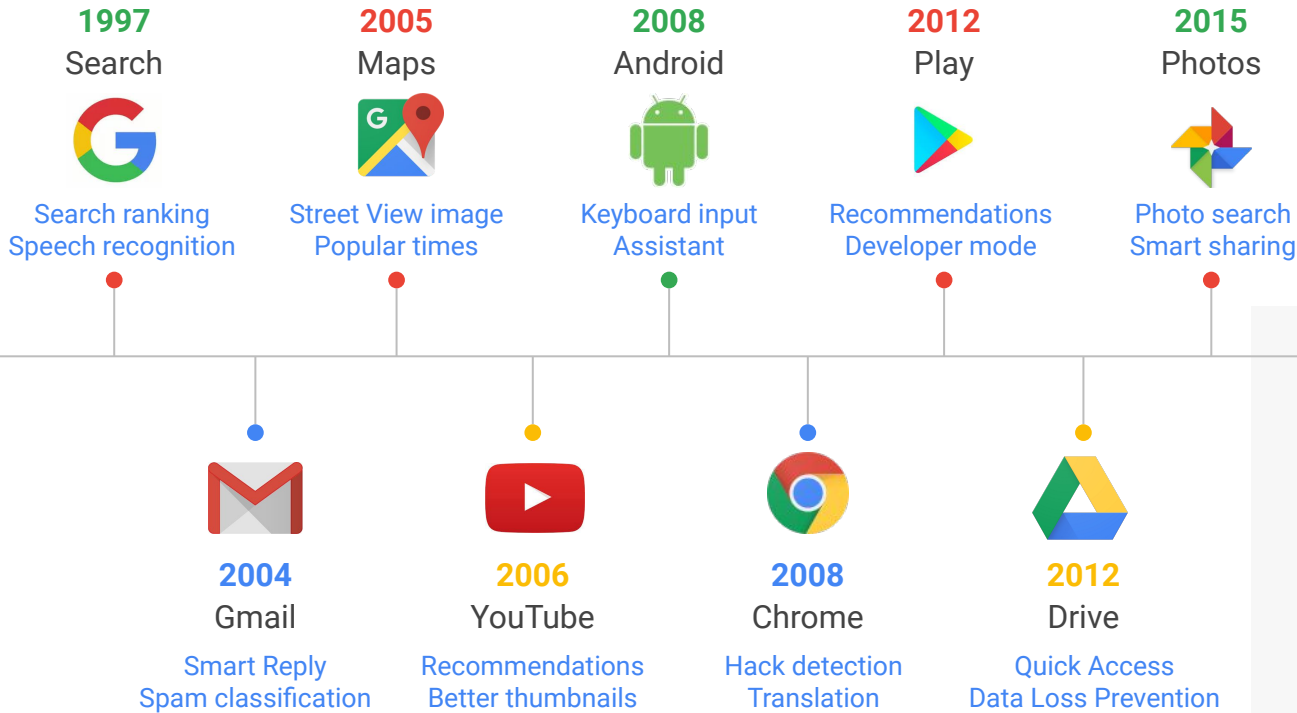
 Sundar Pichai   
@sundarpichai

We want to help businesses and schools impacted by COVID-19 stay connected: starting this week, we'll roll out free access to our advanced Hangouts Meet video-conferencing capabilities through July 1, 2020 to all G Suite customers globally.



Connecting businesses and educators with advanced Hangouts Meet capabilities | Google...  
[cloud.google.com](https://cloud.google.com)

# Our products are **how** we innovate



9

Billion User  
Applications



100%

Machine Learning  
Powered



# Google Cloud

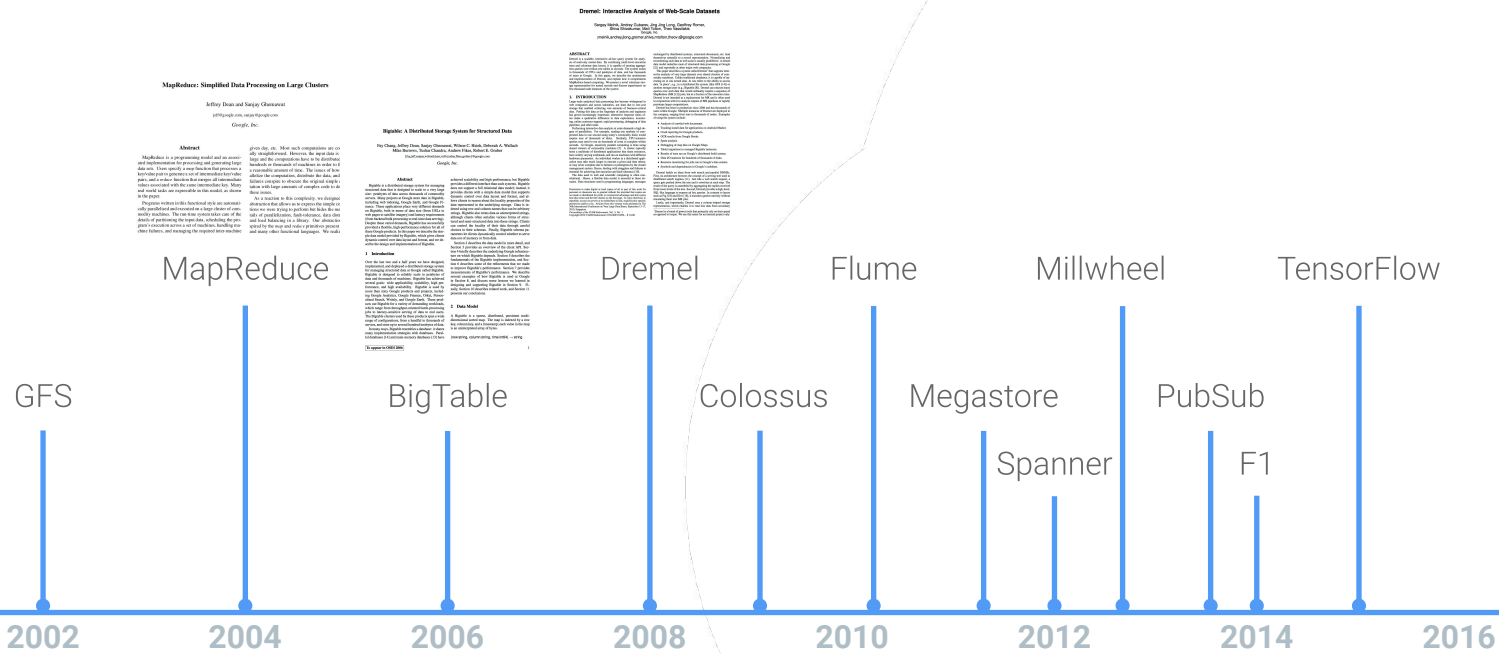
**\$29 billion**  
investment in  
the last 3 years

**Over 1 billion**  
unique IP  
addresses  
served daily

**5,000**  
**caching**  
**points** across  
the globe

Google  
Compute Engine  
guarantees  
**99.95% uptime**

# 15+ years of data research at Google



**MapReduce: Simplified Data Processing on Large Clusters**  
Jeffrey Dean and Sanjay Ghemawat  
jdh@google.com, sghem@google.com  
Google, Inc.

**Abstract**  
MapReduce is a programming model and architecture for processing and generating large data sets. It is a simple, efficient and flexible method for controlling a distributed system (e.g. a cluster of machines) that can be written in Java, and is used to build a wide range of applications. The model is simple enough to be used by non-experts in the field, and is also simple enough to be used by experts in the field.

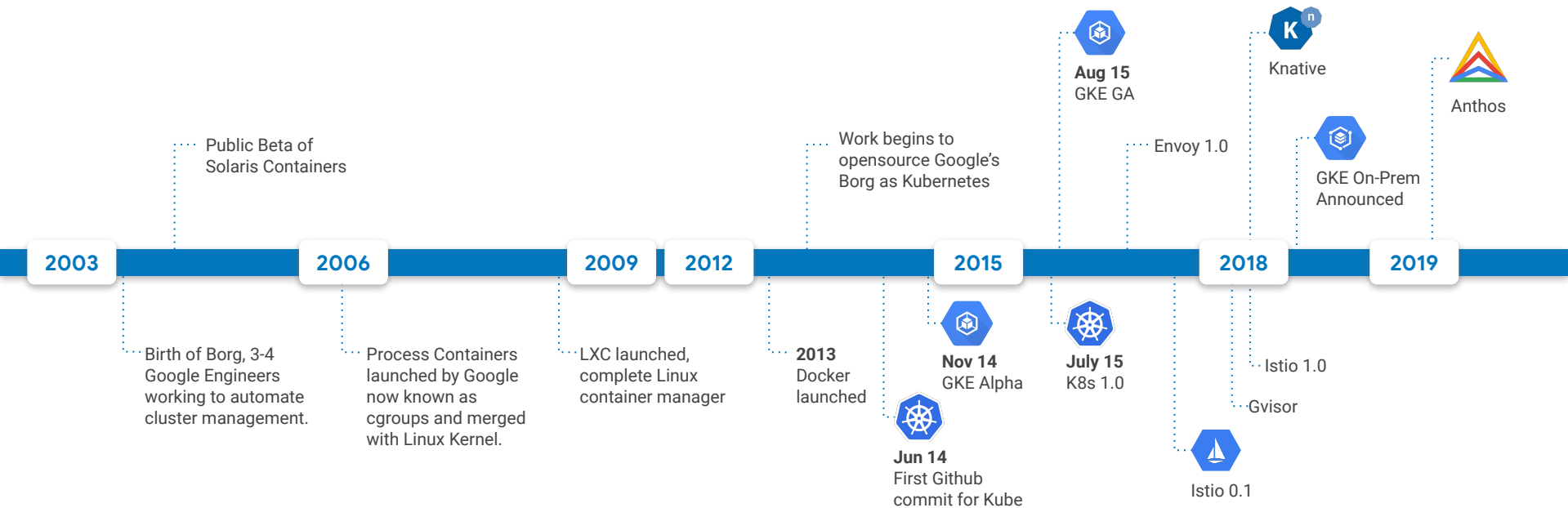
**BigTable: A Distributed Storage System for Structured Data**  
The Chang, Dean, Epstein, Ghemawat, Henkel, Hsieh, Shusterman, Vahanian, White, Zaharia, and Zhou  
chang@google.com, dean@google.com, epstein@google.com, henkel@google.com, hsieh@google.com, shusterman@google.com, vahanian@google.com, white@google.com, zaharia@google.com, zhou@google.com  
Google, Inc.

**Abstract**  
BigTable is a distributed storage system for structured data. It is a simple, efficient and flexible method for controlling a distributed system (e.g. a cluster of machines) that can be written in Java, and is used to build a wide range of applications. The model is simple enough to be used by non-experts in the field, and is also simple enough to be used by experts in the field.

**Dremel: Interactive Analysis of Web-Scale Datasets**  
Hemant Kumar, Sanjay Ghemawat, Rajagopal Srinivasan, and Jeffrey Dean  
hemantk@google.com, sghem@google.com, srinivasan@google.com, dean@google.com

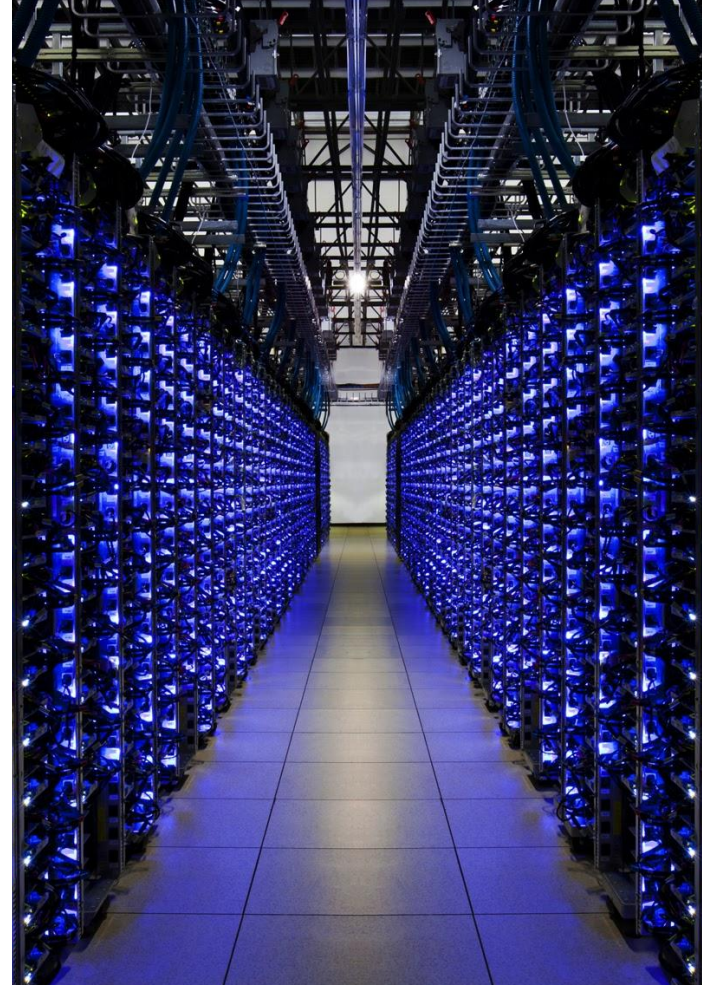
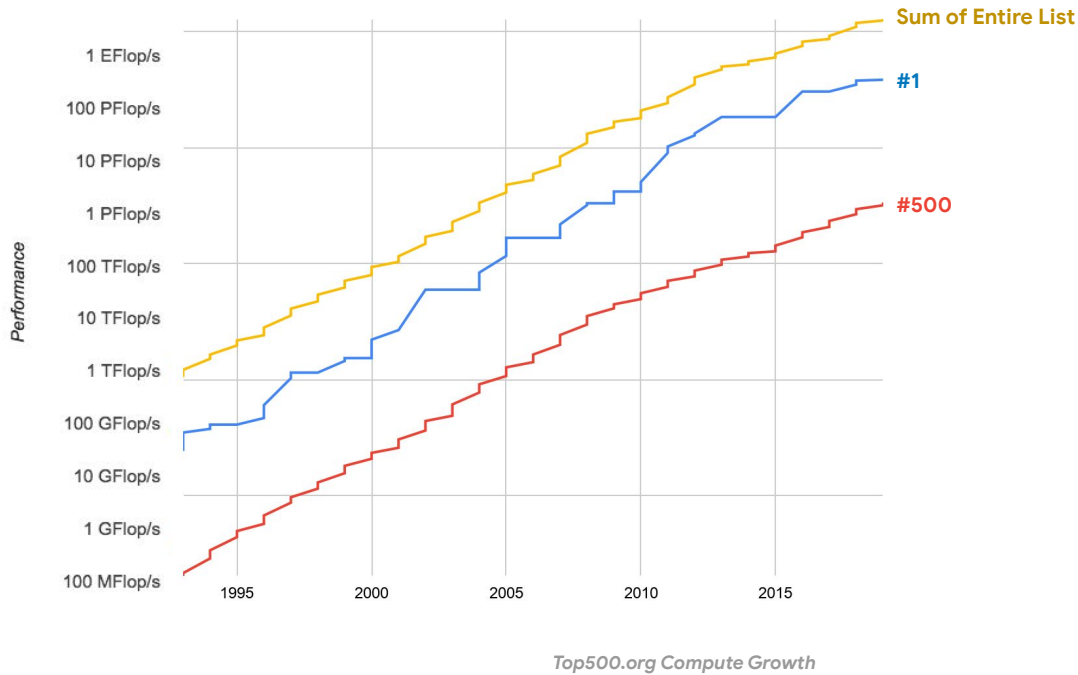
**Abstract**  
Dremel is a distributed system for interactive analysis of web-scale datasets. It is a simple, efficient and flexible method for controlling a distributed system (e.g. a cluster of machines) that can be written in Java, and is used to build a wide range of applications. The model is simple enough to be used by non-experts in the field, and is also simple enough to be used by experts in the field.

# 15+ years experience in application orchestration



**Today, on average, we launch 4 billion containers per week**

# Compute needs are growing



# Data is now BIGGER than big

2008

Los Alamos National  
Laboratory - RoadRunner

3PB

Capacity - Panasas Storage

55 GB/s  
throughput

2018

Oak Ridge National Laboratory  
- Summit

250PB

Capacity - IBM Spectrum Scale

2.5 TB/s  
throughput

83x

45x



Democratize **high performance computing**  
and make it universally accessible and useful.

# We continue to make investments in HPC



Auto-scaling and hybrid features built for Google Cloud

Mar 2019



DDN releases Cloud Edition for Lustre on GCP, places #8 on IO500

Apr 2019



High-performance, scalable file storage, acquired by Google

Aug 2019



Popular HPC platform providing a GUI to run HPC workloads on GCP

Nov 2019

H1 2019

H2 2019

Apr 2019



The first major provider to offer T4 GPU globally

Apr 2019



Computed Optimized VMs (C2) announced. 40% higher performance than N1 VMs

Aug 2019



EPYC VMs (N2D) announced. Up to 60% better platform memory bandwidth than existing instances

Nov 2019



Batch on GKE

Cloud-native workload manager for containerized HPC batch jobs

# Outcomes

Clemson broke the world record for running an HPC workload using the most compute cores on any commercial cloud.

**2.138M vCPU | 133,573 instances | 200+ TB data processed**

Average cost of **\$0.008 USD per vCPU hour**

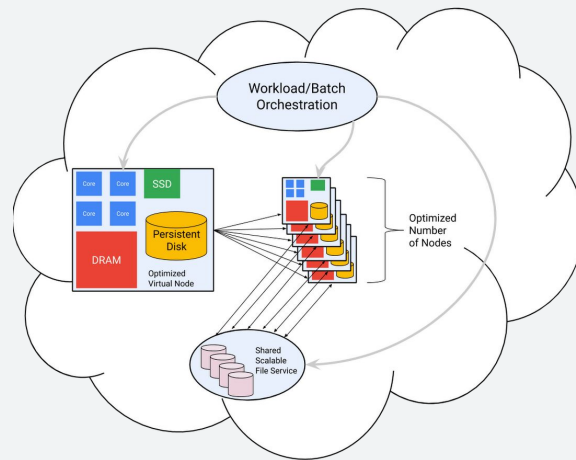
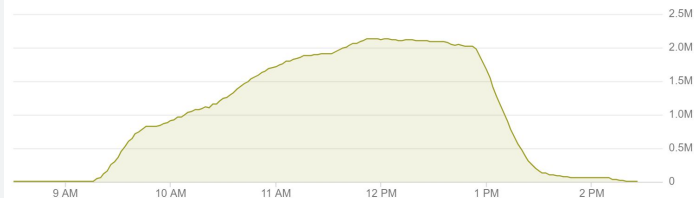
Published **“On-Demand Disaster Management using High Performance Computing in the Commercial Cloud,”** in The Journal of Supercomputing

A corpus of traffic data for public use to be added to the GCP Marketplace via the Public Dataset Program

[Google Cloud Blog Post](#)



GCP CPU Core Ramp and Count





# Computing

# Google Compute Engine

## Compute power

- Up to 416 vCPUs
- Up to 12TB RAM
- Up to 3TB Local SSD
- Up to 8 GPUs
- Up to 7TB Intel Optane

## Networking

- Globally routed subnets default
- 2Gbps per vCPU, up to 32Gbps
- 15,000 VMs per Virtual Private Cloud (VPC)

## Custom Machine Types

- Choose your Core:RAM ratio
- Recommendation engine for optimum machine size

## Compute Optimized VMs

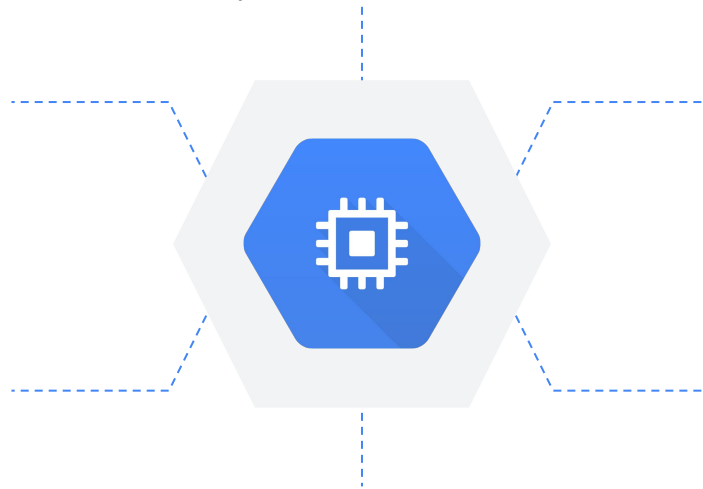
- Fast Clocks - 3.8 GHz
- Up to 60 vCPU, 240GB RAM
- vNUMA exposure
- Overhead reduction

## Accelerators

- Up to 8 NVIDIA GPUs
- NVIDIA K80, P100, P4, V100, T4
- NVLink and PCIe x16 for maximum performance
- Google Cloud TPUs for ML

## Preemptible VMs

- Up to 80% discount
- Custom Machine Types, GPUs, Local SSDs



# NVIDIA GPUs



Attach up to eight GPUs per instance



NVLink and PCIe x16 to achieve maximum performance



Per-second billing, Sustained Use Discounts  
Preemptible support (~70% off)



NVIDIA Tesla K80  
NVIDIA Tesla P100  
NVIDIA Tesla P4  
NVIDIA Tesla V100  
NVIDIA Tesla T4



# “The Jungle Book” wins Best **VFX** Oscar in 2017.

Rendered on Google Cloud Platform.

**5.5 million** core hours

**41 years** of compute time over two months

**1.5 million** tasks processed

(MPC film, 2017)



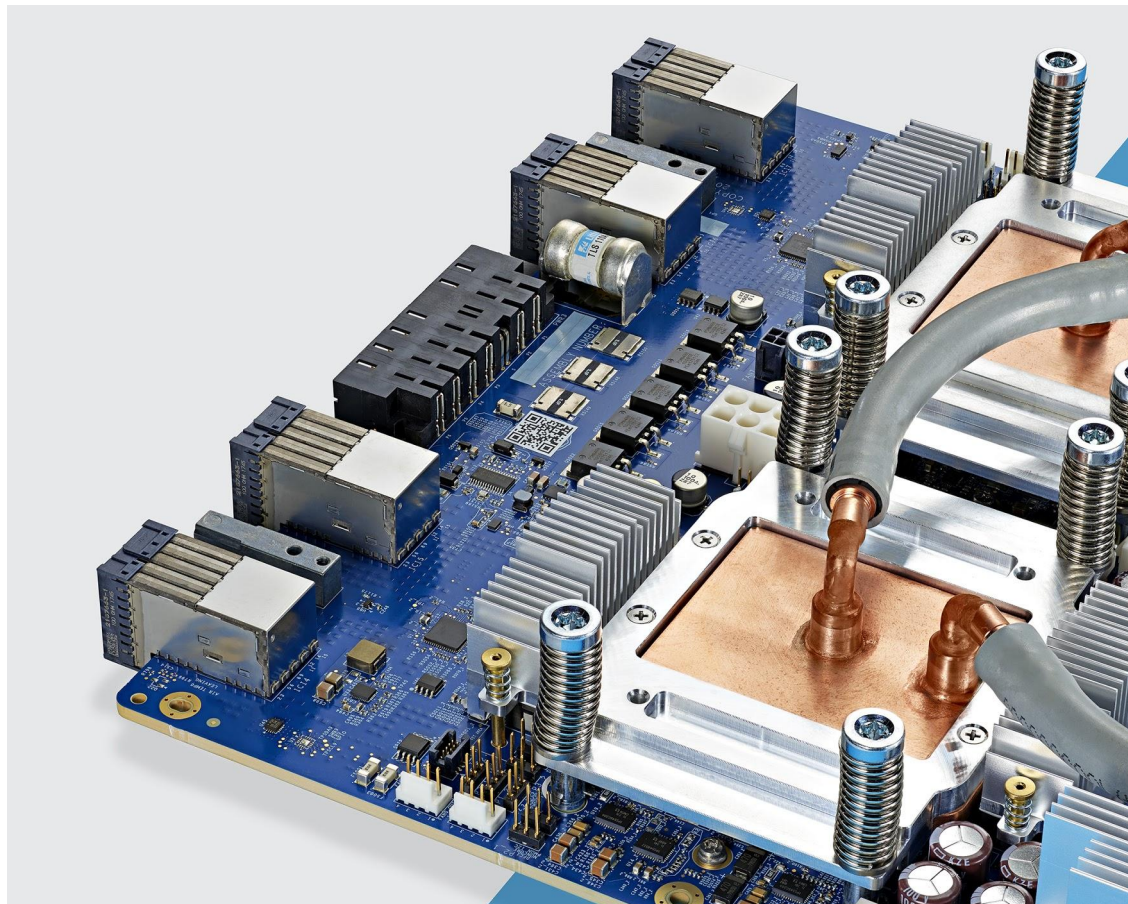
# Cloud TPU v3

420 teraflops

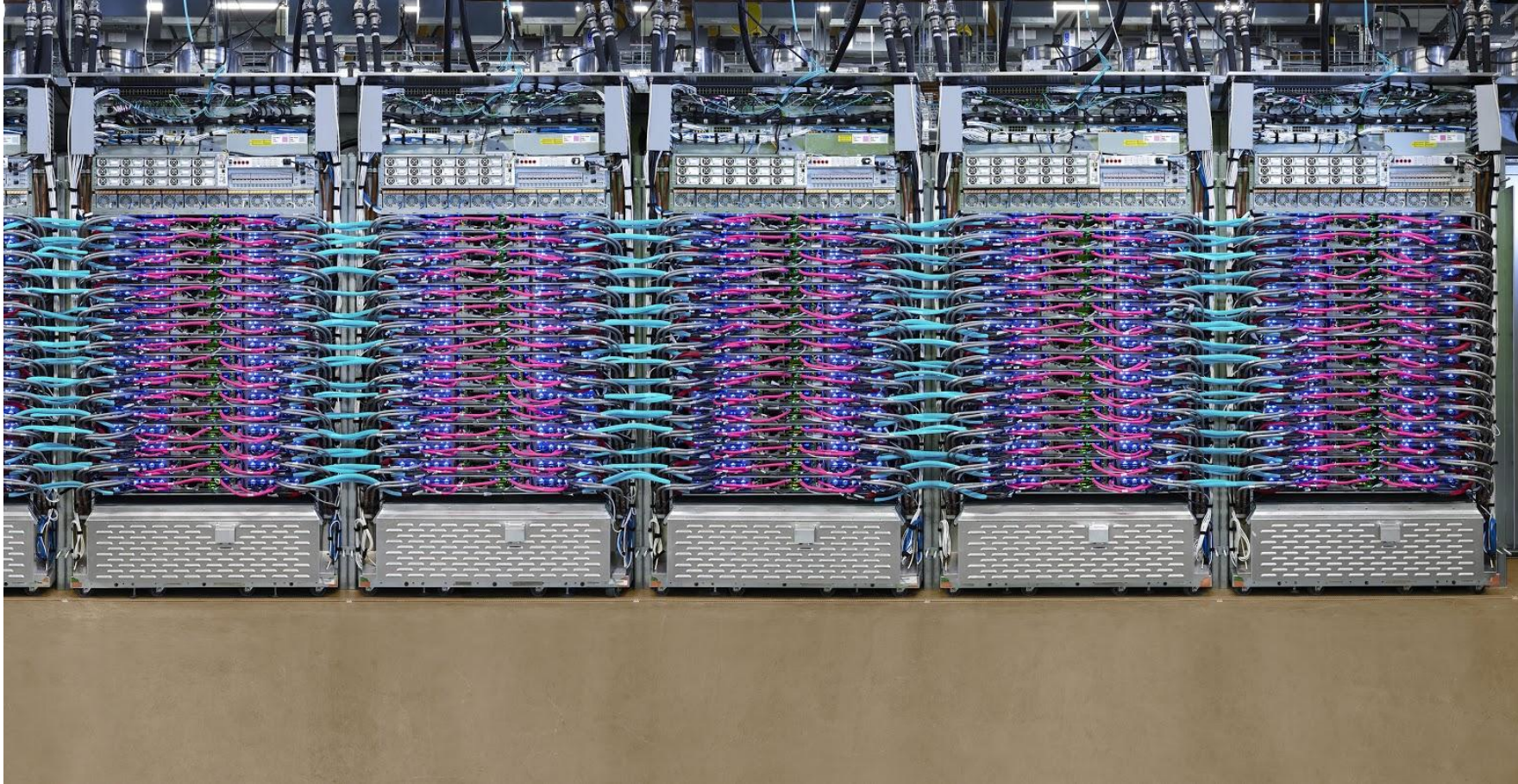
128 GB of high-bandwidth memory

Training and inference

Generally Available (GA)





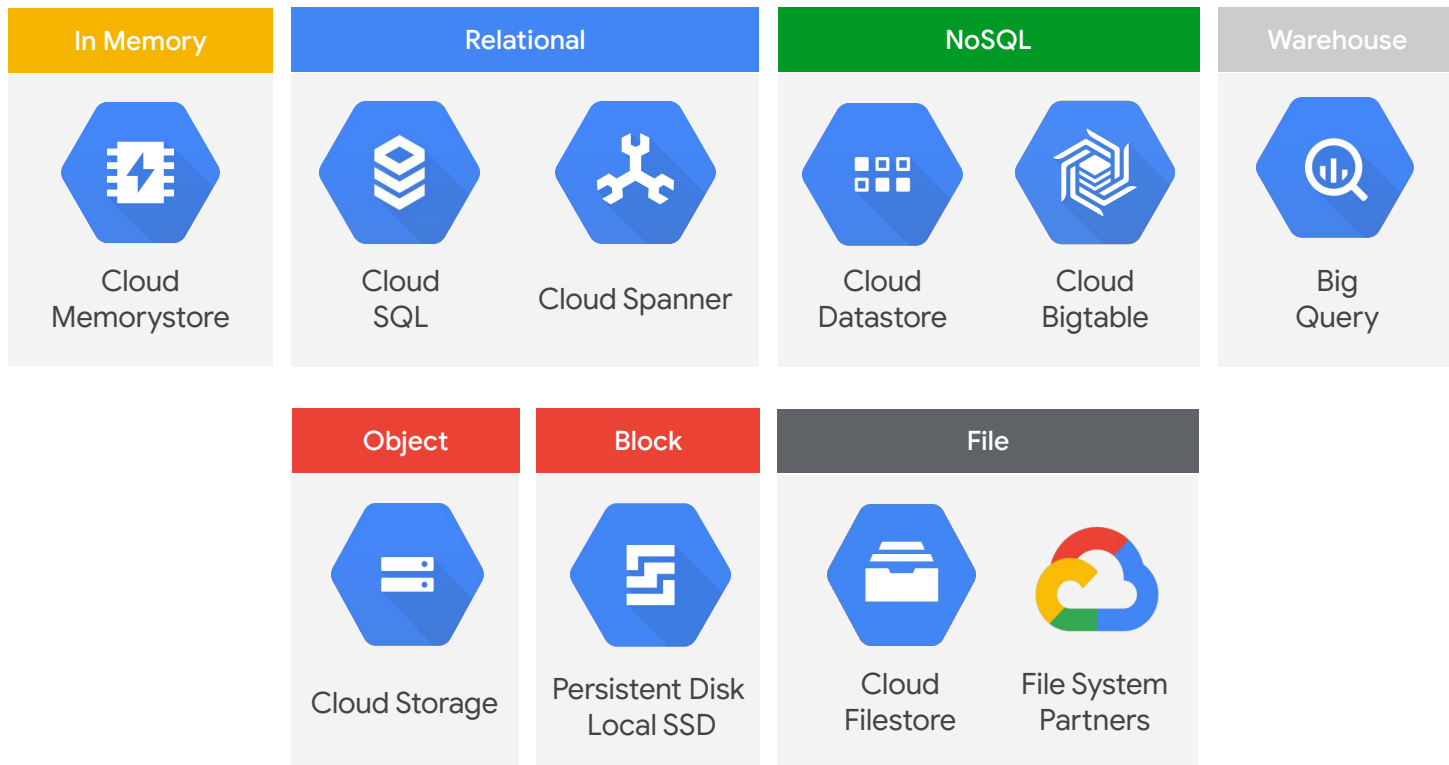


Cloud TPU v3 Pod: >100 PFLOPS  
(8x faster than v2 Pod)



Storage

# Where do I store my data in Google Cloud?



# HPC Storage on GCP



## Google Cloud Storage

Exabyte-scale, feature-rich object storage  
Automatically scaling throughput



## Persistent Disk

SSD/HDD Persistent Disk  
High-performance, replicated block storage



## Local Storage

Local SSD (NVMe) for scratch and fast access  
Physically attached to node via PCI



## Cloud Filestore

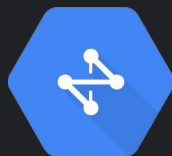
Highly available, durable, POSIX-compliant shared storage  
across tens of thousands of nodes



## Partner, hybrid, and open-source

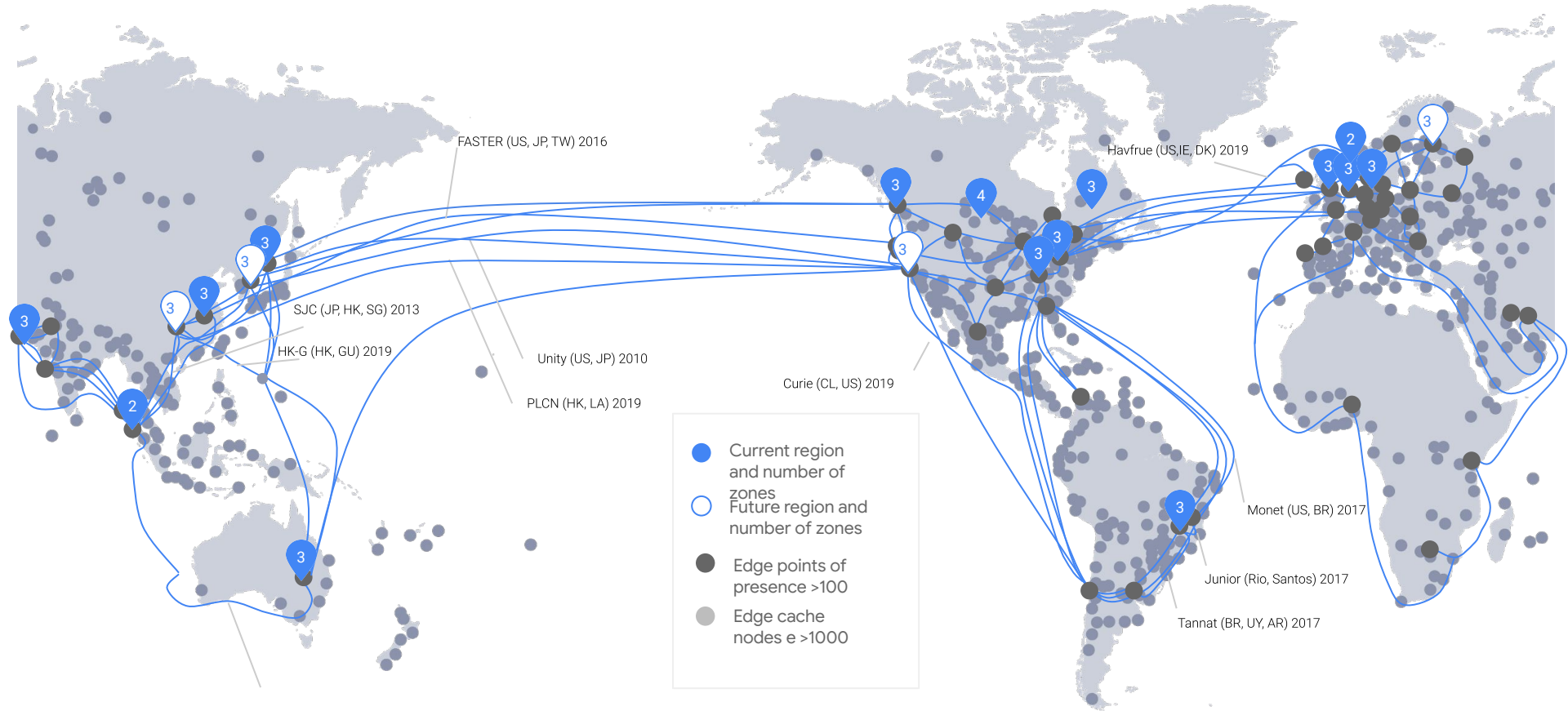
Storage solutions for NetApp, Elastifile, DDN, Lustre, and more  
Move petabytes to GCS with the Data Transfer Appliance





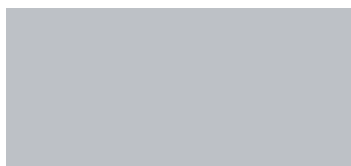
# Networking

# Global network infrastructure



# Bisectional bandwidth

200 Tb/sec  
entire internet



1,000+ Tb/sec  
Single Google  
data center



# Google Networking

## Performance

- 15,000 VMs per VPC
- Predictable, low latency (~20 - 40  $\mu$ s)
- Scalable bandwidth
  - 2 Gbps per vCPU
  - Up to **32Gbps** per VM
- Open-sourcing and tailoring latency-sensitive tools to GCP (gRPC)

## Global Network

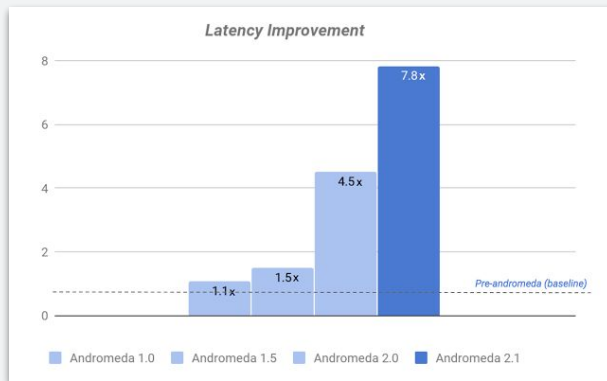
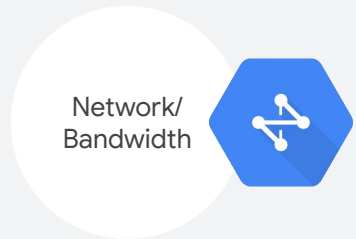
- Thousands of POP around the world
- Google Backbone between datacenters
- Multiple interconnect options to on-prem



## Network

### Efficient SDN network

- Clos topology, a collection of smaller custom switches arranged to provide the properties of a much larger logical switch.
- Centralized software management stack.
- Relying more on custom protocols tailored to the high performance data center.





# Batch on GKE *Preview*

Batch Job Management Platform built for Kubernetes; available on GKE

- Extends K8S APIs to add support for Batch Jobs and Job Queues.
- New Autoscaler and Scheduler optimized for batch workloads



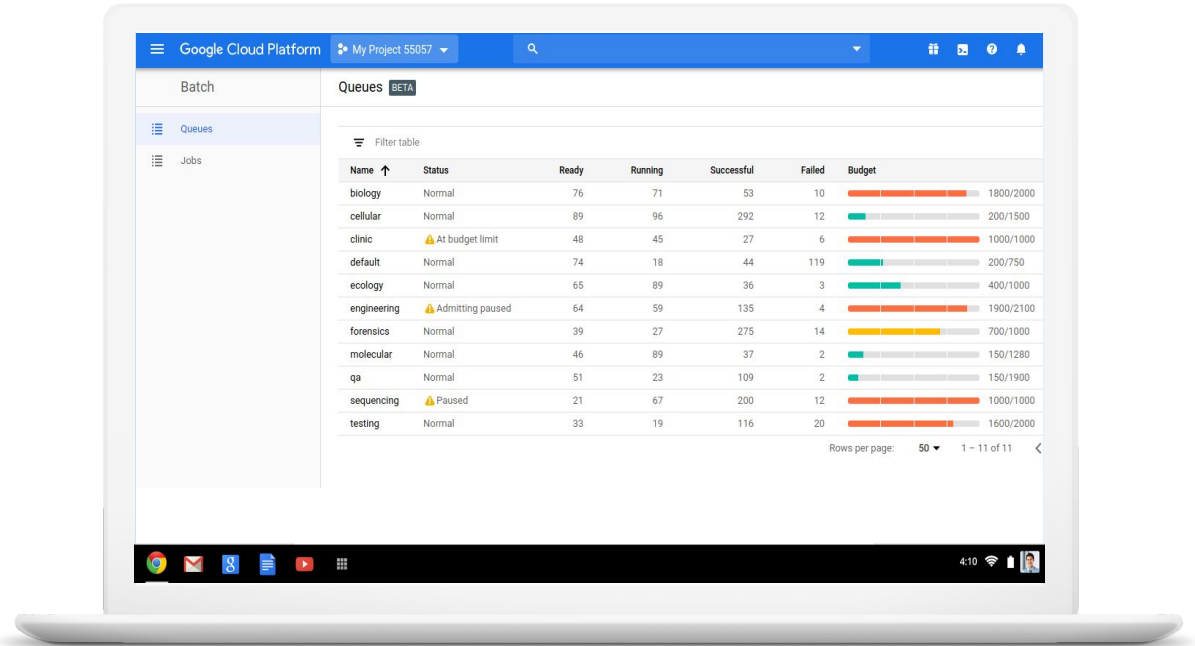
# Features

## Batch Constructs

1. Queues and Queue priorities
2. Job dependencies and constraints
3. Array Jobs

## Cost Management

4. Budget Controls
5. Autoscaling
6. Rightsizing



# Features (Contd)

## Ease of use

1. User account management
2. Data copy to/from cluster
3. Ksub command line tool for easy job submissions and management

## Platform Integration

4. GPU support
5. Stackdriver Monitoring
6. Persistent job logs



The screenshot shows the Google Cloud Platform interface for a job detail. The top navigation bar includes the Google Cloud Platform logo, a 'Select a project' dropdown, and a search icon. The main content area is titled 'Job Detail' with a 'BETA' badge. On the left, there is a sidebar with 'Queues' and 'Jobs' options. The main panel displays the job's status as 'Succeeded' with a green checkmark, and the submission date as '2 days ago'. Below this, there are tabs for 'DETAILS' and 'TASK LOGS'. The 'Job information' section includes a table with the following data:

Name	clinic-run1
Submission date	2 days ago
Last status change	2 days ago
Tasks complete	10 of 10
Cost	120

The 'Queue info' section includes a table with the following data:

Name	clinic
Priority	high
Admitting	0 of 3
Scheduling	No
Jobs ahead	20

At the bottom, there are three budget utilization bars:

- Budget window:** 3 of 7 days, with a bar showing approximately 30% utilization of a 7-day window.
- Daily limit:** 400/600, with a bar showing approximately 67% utilization.
- Job limit:** 300, with a bar showing approximately 33% utilization.



**Thank you**

Google Cloud