Get Your Head in the Cloud: Lessons in GPU Computing with Google Cloud & Schlumberger

Wyatt Gorman - HPC Specialist, Google Cloud
Abhishek Gupta - HPC/Viz Innovation Lead, Schlumberger STIC

Google Cloud

Our Speakers

Wyatt Gorman
HPC Specialist,
Google Cloud

Abhishek Gupta
HPC Tech Lead,
Schlumberger

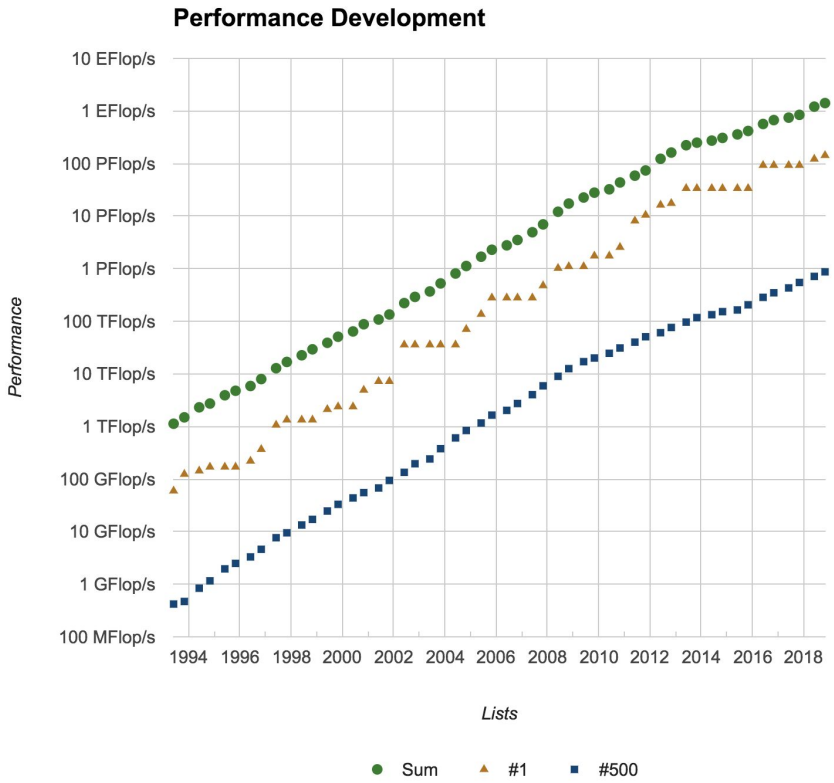
Google Cloud

HPC in Google Cloud

Google Cloud

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

Compute needs are growing



Google Cloud

https://www.top500.org/statistics/perfdevel/

Data is now bigger than big

Ten years in HPC storage

2008

Los Alamos National Laboratory - RoadRunner

3PB capacity Panasas Storage

55 GB per second throughput

83x

45x

2018

Oak Ridge National Laboratory - Summit

250PB capacity of IBM Spectrum Scale

2.5 TB per second throughput

Google Cloud

#8 storage system in the IO-500.

Build your own version with DDN's GCP Marketplace solution (currently in EAP).

IO500

#	information							io500		
	institution	system	storage vendor	filesystem type	client nodes	client total procs	data	score	bw	md
									GiB/s	kIOP/s
1	Oak Ridge National Laboratory	Summit	IBM	Spectrum Scale	504	1008	zip	366.47	88.20	1522.69
2	Korea Institute of Science and Technology Information (KISTI)	NURION	DDN	IME	2048	4096	zip	160.67	554.23	46.58
3	University of Cambridge	Data Accelerator	Dell EMC	Lustre	528	4224	zip	158.71	71.40	352.75
4	JCAHPC	Oakforest-PACS	DDN	IME	2048	16384	zip	137.78	560.10	33.89
5	WekaIO	WekaIO	WekaIO		17	935	zip	92.95	37.39	231.05
6	KAUST	ShaheenII	Cray	DataWarp	1024	8192	zip	77.37	496.81	12.05
7	University of Cambridge	Data Accelerator	Dell EMC	BeeGFS	184	5888	zip	74.58	58.81	94.57
8	Google	Exascaler on GCP	Google	Lustre	120	960	zip	56.77	23.06	139.74
9	JCAHPC	Oakforest-PACS	DDN	Lustre	256	8192	zip	42.18	20.04	88.78
10	KAUST	ShaheenII	Cray	Lustre	1000	16000		41.00*	54.17	31.03*

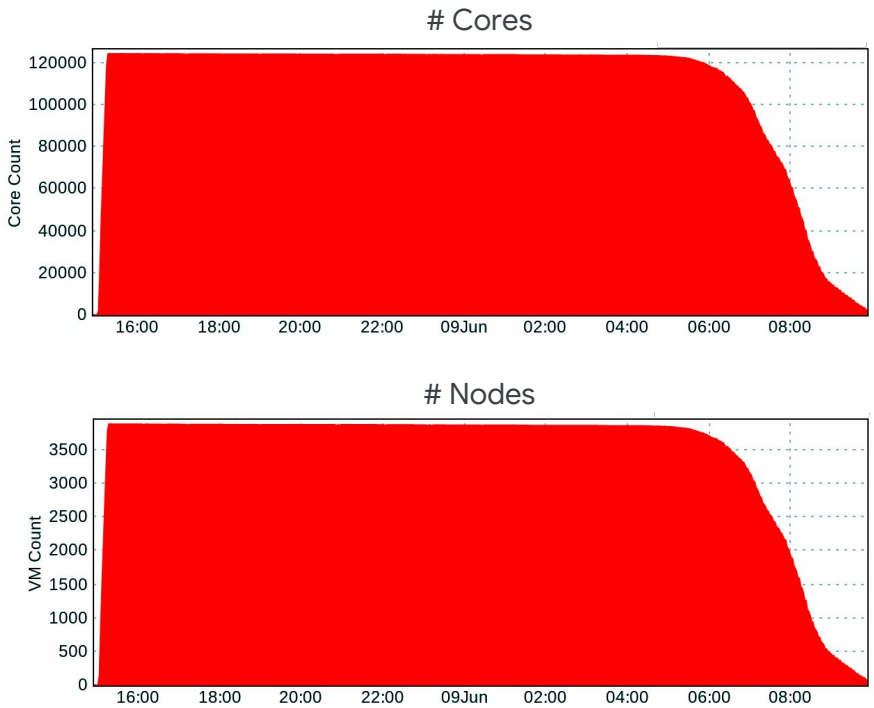
Google Cloud

University of South Carolina

Microbiome sequencing

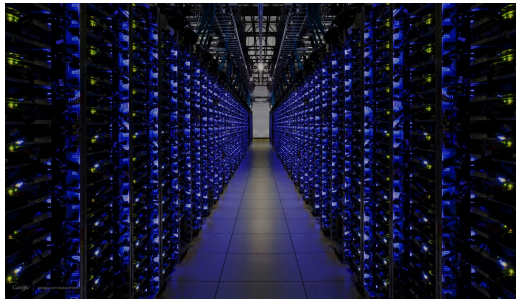
- 4000 Nodes
- 125,000 Cores
- 16 hrs of execution
- Slurm Auto-Scaling Cluster
  - Single job
- Preemptible VMs
- R/O Persistent Disk/GCS

https://edu.google.com/latest-news/case-studies/sc-gcp



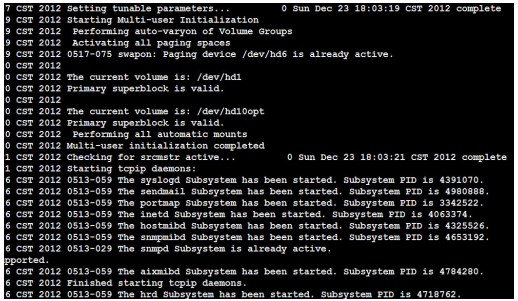
Google Cloud

Outline



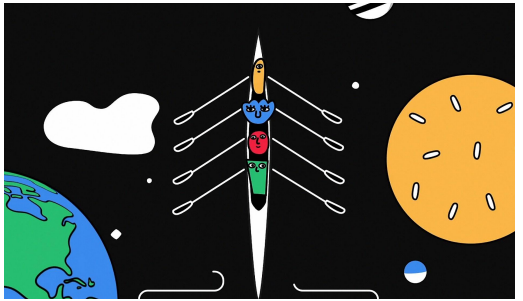
HPC Infrastructure

HPC is entering a new era defined by new accelerator technologies, rapidly enabling significant new advances.



HPC Software

Google's approach is towards open-source licensing, and partnering with popular vendors.



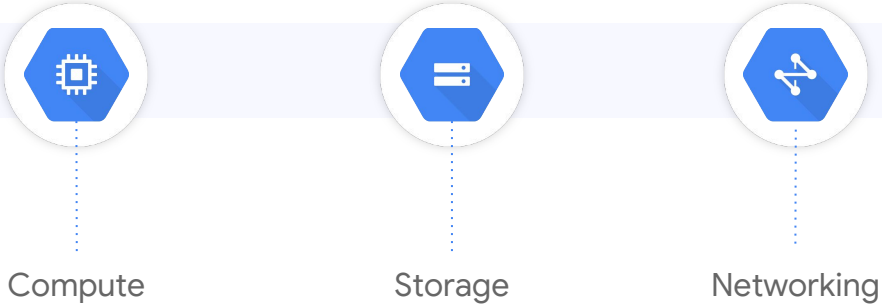
Community and Partnerships

Building new communities and engaging existing ones to partner together to democratize HPC.

Google Cloud

HPC Infrastructure

HPC Infrastructure



Google Cloud

Google Compute Engine

Highly configurable resources

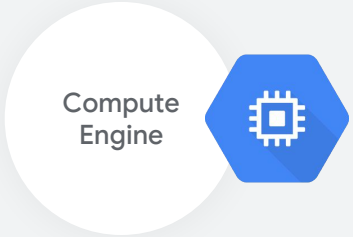
- Latest high-performance processors
- Customizable instances
  - 1 to 160 Cores
  - Up to 3844 GB RAM
  - Up to 7 TB Intel Optane DRAM
- Preemptible VMs
- NVIDIA GPUs

Managed compute paradigms

- Pipelines API
- Managed instance groups
- Deployment Manager
- Cloud Composer (Airflow)

Supports various workloads

- MPI
- Batch/HTC
- Real time
- MapReduce
- Machine learning



Works with your favorite scheduler

- Slurm
- HTCondor
- LSF
- Grid Engine
- Microsoft HPC Pack
- More coming soon...

Google Cloud

NVIDIA GPUs

Attach up to eight GPUs per instance

Attached directly to the VM via PCIe x16
to achieve near bare-metal 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

Google Cloud

Preemptible VMs

Made for batch, checkpointed, and high throughput computing.

Super-low-cost, short-term instances

- Up to 80% cheaper than standard instances[1]
- Preemptible pricing for GPUs (~70% off), Cloud TPUs (~70% off), and Local SSDs (~40% off)
- Maximum lifetime of 24 hours, may be preempted with 30-seconds notice
- Simple to use – preemptible flag on any instance type[1]
- Handle cleanup with startup and shutdown scripts

Ideal for a variety of workloads

- Genomics, pharmaceuticals
- Physics, math, computational chemistry
- Data processing (for example, with Hadoop or Cloud DataProc)
- Image handling, rendering, and media transcoding
- Monte Carlo simulations
- Financial services

[1]https://cloud.google.com/preemptible-vms/

Google Cloud

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

5.5 million core hours
41 years of compute time over two months
1.5 million tasks processed

(MPC film, 2017)

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

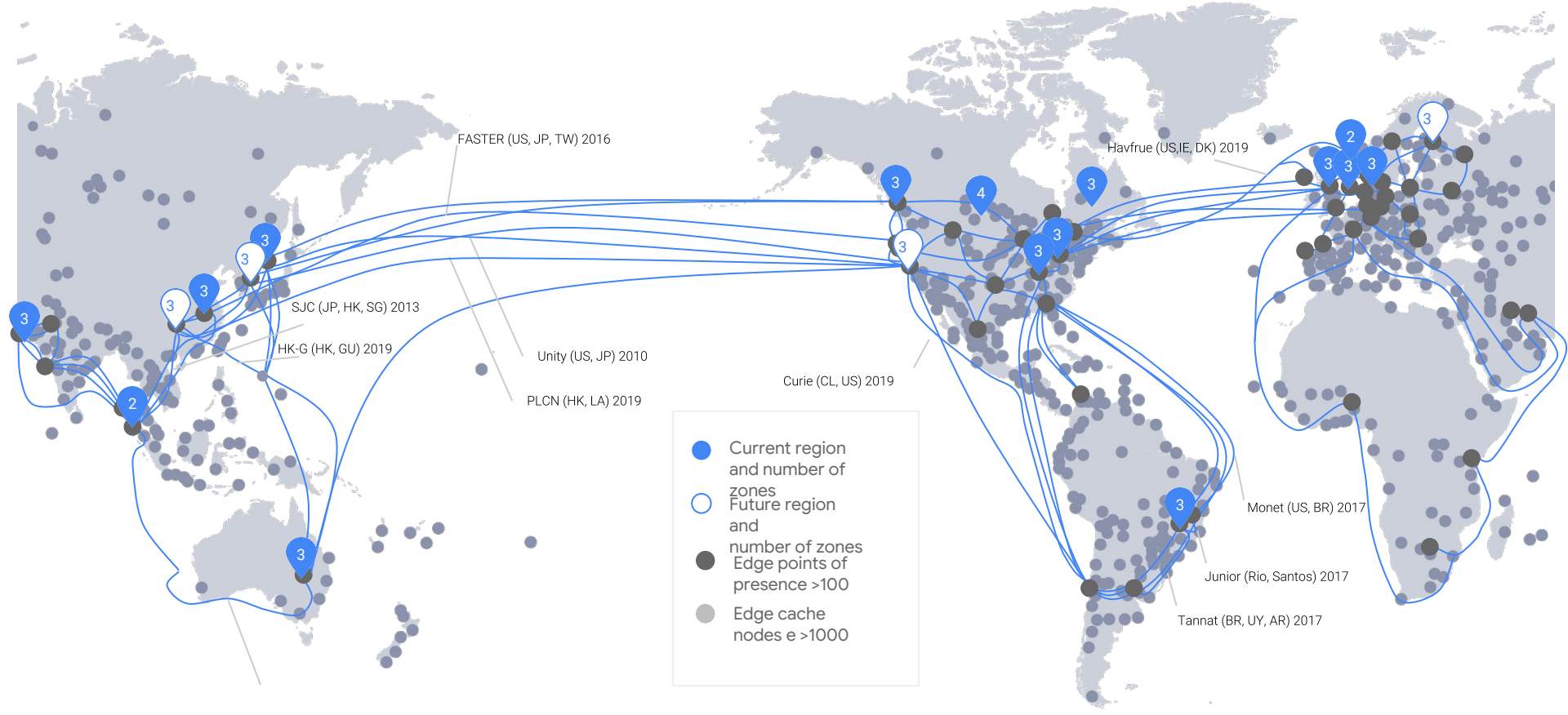
Local Storage
Local SSD (NVMe) for scratch and fast access
Physically attached to node via PCI

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

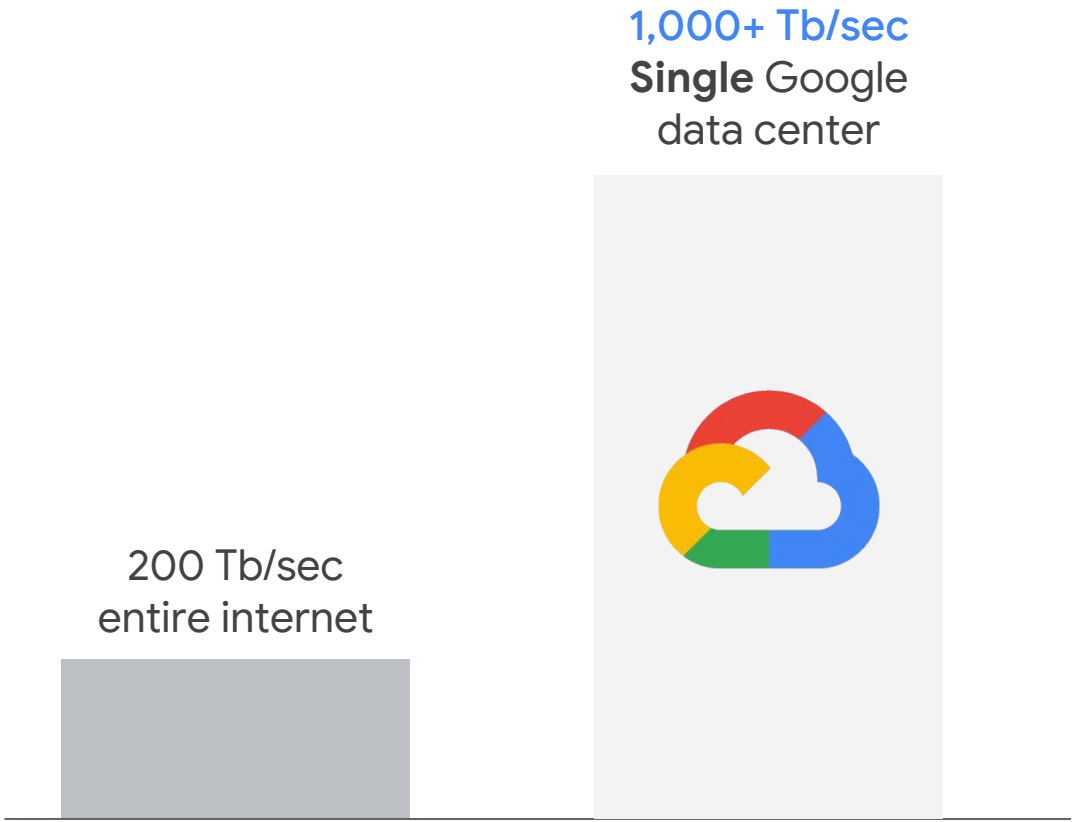
Google Cloud

Global network infrastructure



Google Cloud

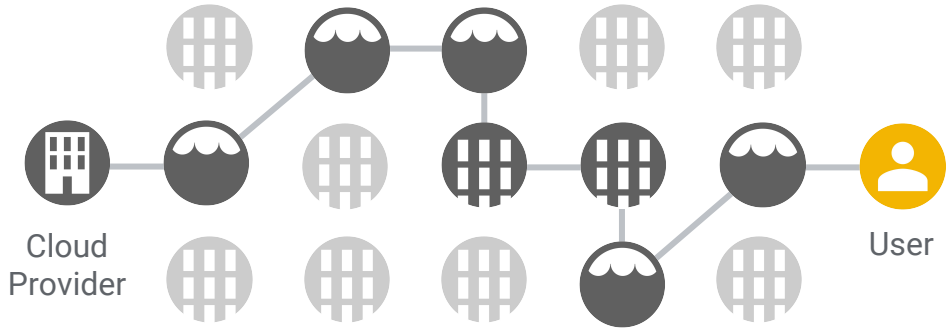
Bisectional bandwidth



Google Cloud

https://cloudplatform.googleblog.com/2015/06/A-Look-Inside-Googles-Data-Center-Networks.html

The network matters.



Typical cloud provider

Google Cloud

Google Cloud

Google Networking

Performance

- 7,000 VMs per VPC
- Predictable, low latency (~20 - 40 µs)
- Scalable bandwidth
  - 2 Gbps per vCPU
  - Up to **16 Gbps** per VM
- Open-Sourcing high-performance internal protocols and tools (gRPC)
- Tailoring latency-sensitive tools to GCP

Global Network

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

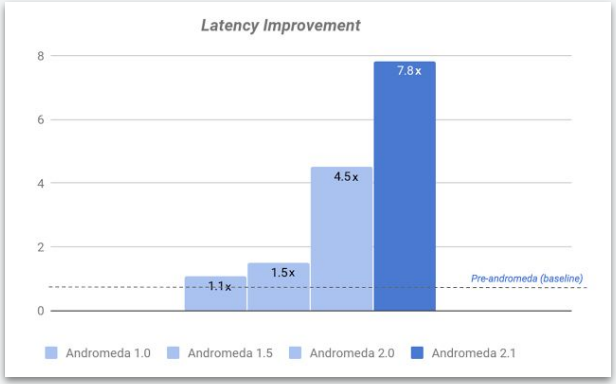
Google Cloud

Network

Network/ Bandwidth

Efficient SDN network

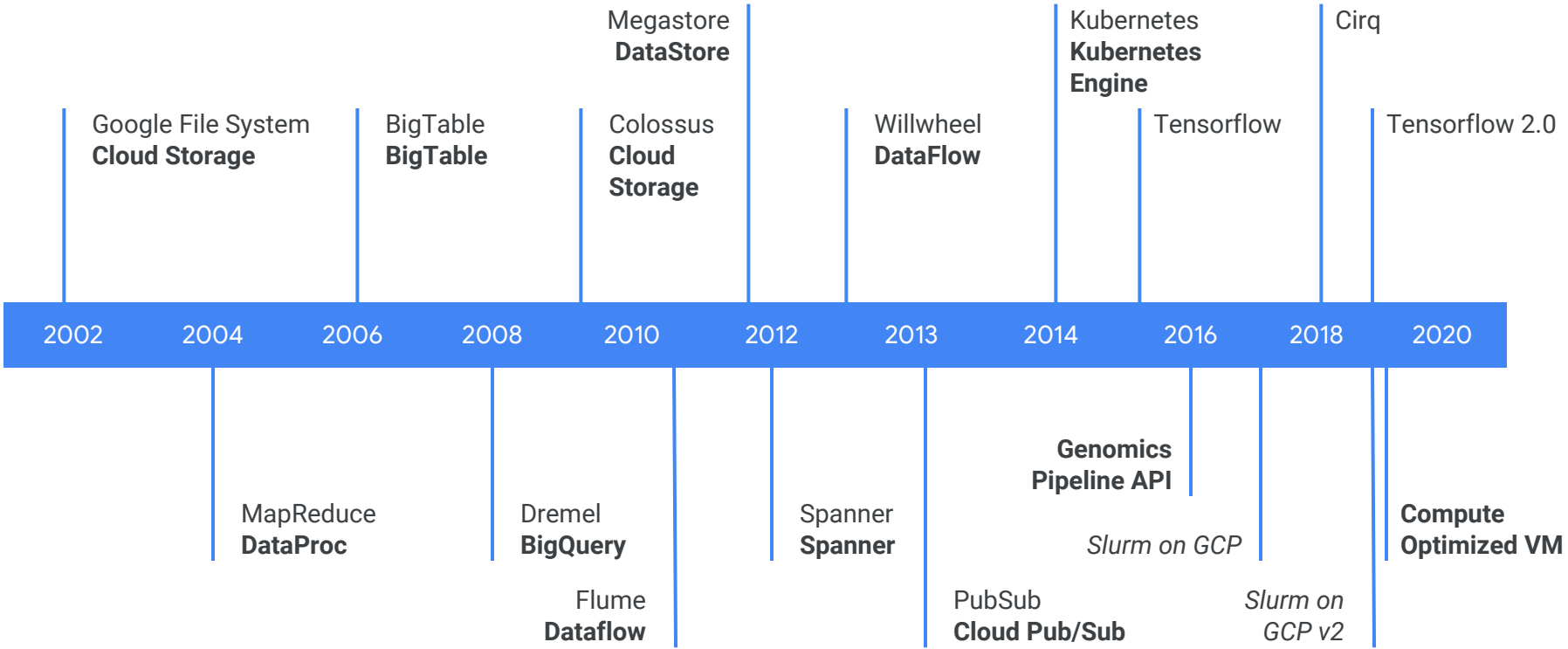
- Clos topology, a collection of smaller custom switches arranged to provide the properties of a much larger logical switch.
- Centralized software management stack.
- Relying more on custom protocols tailored to the high performance data center.



HPC Software

Google advances scientific computing

Research
Cloud Solution
Partner Solution



Google Cloud

Growing ecosystem of HPC partners

Storage

High performance storage systems for HPC workloads

DDN STORAGE · DELL EMC · elastifile · Quobyte · pixitmedia · NetApp

Infrastructure and Workload Managers

Tools and platforms for job scheduling and cluster management

Workload Managers

slurm workload manager · HTCondor High Throughput Computing · Adaptive COMPUTING

HPC as a Service

[:] · Parallel Works · PBScloud.io

Infrastructure

intel · NVIDIA

HPC applications

HPC powered applications targeting vertical and horizontal use cases

Life Sciences

BROAD INSTITUTE

Manufacturing

Altair

Interactive Computing

TECHILA

Electronics Design Automation

SYNOPSYS® · metrics

SI Partners

Boutique SI

FluidNumerics

Global SI

appsbroker · intel

Future

UNIVA · Altair · alcesflight · rescale · Omnibond

ANSYS

Atos

Google Cloud

Slurm Workload Manager

Google partnered with SchedMD to integrate the Slurm Workload Manager with GCP to harness the elasticity of Compute Engine

Three ways to use Slurm:

- **Cloud Auto-Scaling:** Automatic elastic scaling of instances, on demand, according to queue depth and job requirements. Spins resources down once idle timeout is reached.
- **Burst to Cloud:** Dynamically create virtual machines to offload jobs from your on-premise cluster to Google Cloud. Leverages Cloud Auto-Scaling functionality.
- **Federate to Cloud:** Federate jobs between your on-premise Slurm cluster and your Google Cloud Slurm cluster(s).
- **Open Source on SchedMD's Github:** https://github.com/schedmd/slurm-gcp

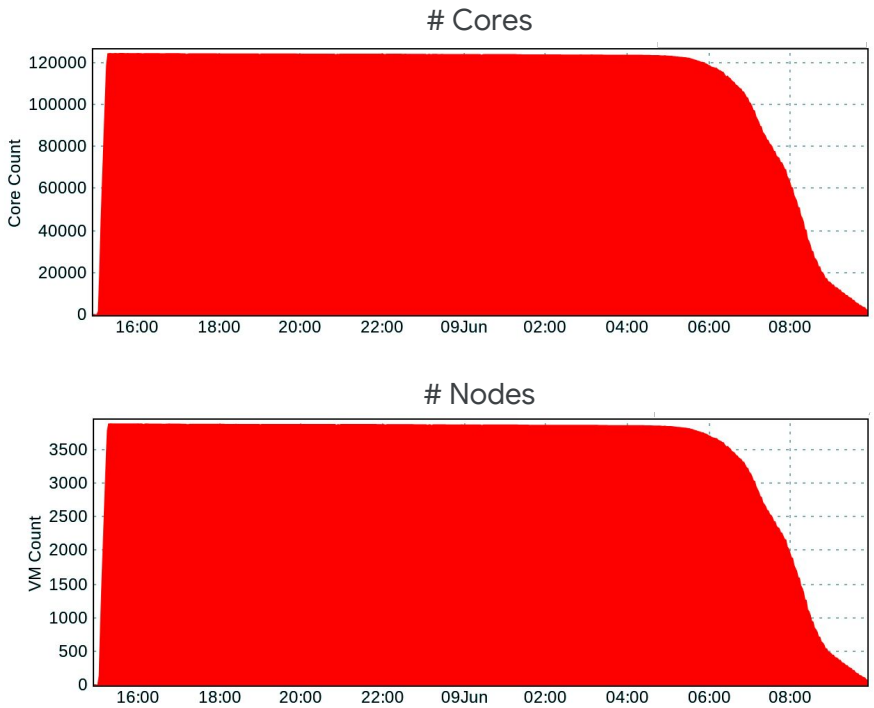
Google Cloud

University of South Carolina

Microbiome sequencing

- 4000 Nodes
- 125,000 Cores
- 16 hrs of execution
- Slurm Auto-Scaling Cluster
  - Single job
- Preemptible VMs
- R/O Persistent Disk/GCS

https://edu.google.com/latest-news/case-studies/sc-gcp



Google Cloud

Community support

El Romero solar farm in Chile (80 MW for Google)

RENEWABLE ENERGY: BY THE NUMBERS

3 GW
of renewable energy

Google is the world's largest corporate purchaser of renewable energy. We've signed 26 agreements totaling nearly 3 GW of renewable energy—generating emissions savings equivalent to taking more than 1.3 million cars off the road per year.

100%
renewable energy

In 2017, we matched 100% of the electricity consumption of our operations with purchases of renewable energy.

$2.5 billion
in investment commitments

Since 2010, we've committed to invest nearly $2.5 billion in renewable energy projects with a total combined capacity of 3.7 GW.

11 years
of carbon neutrality

Google has been carbon neutral since 2007. Because of our renewable energy and carbon offset programs, our net operational carbon emissions during this period were zero.

GCP Credits for Events, Courses, and Research

Cloud HPC Days and Credits for Workshops
Lab-based HPC Days with customer panels, partner demos, and community networking. Cloud credits for hackathons and workshops.

GCP Research Credits and Faculty Grants
Free credits for academic research workloads and for student learning & coursework

Google Cloud

Google and the National Science Foundation

Google Cloud partnered with NSF to simplify accessing cloud computing resources:

- BigData Solicitation (2017 & 2018 only)
- **[NEW]** Cloud Access
- **[NEW]** CC* Solicitation
- **[NEW]** NSF/Internet2 Cloud Pilot
- **[NEW]** NSF-ML/GCP - between University Relations with Material Sciences

Google Cloud

Schlumberger

02

Google Cloud

Seismic HPC and visualization with GPUs in Google Cloud - the innovation team

Anthony Lichnewsky
HPC architect
Schlumberger

Carlos Boneti
HPC Tech Lead, Energy
Google Cloud

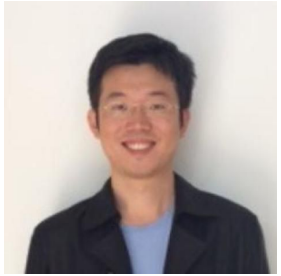
Christopher Leader
Software geophysicist
Schlumberger

Kanai Pathak
Solutions Lead, Energy
Google Cloud

Abhishek Gupta
HPC Tech Lead
Schlumberger

Andrei Khlopotine
Senior HPC engineer
Schlumberger

Markus Schlafli
Senior Viz engineer
Schlumberger

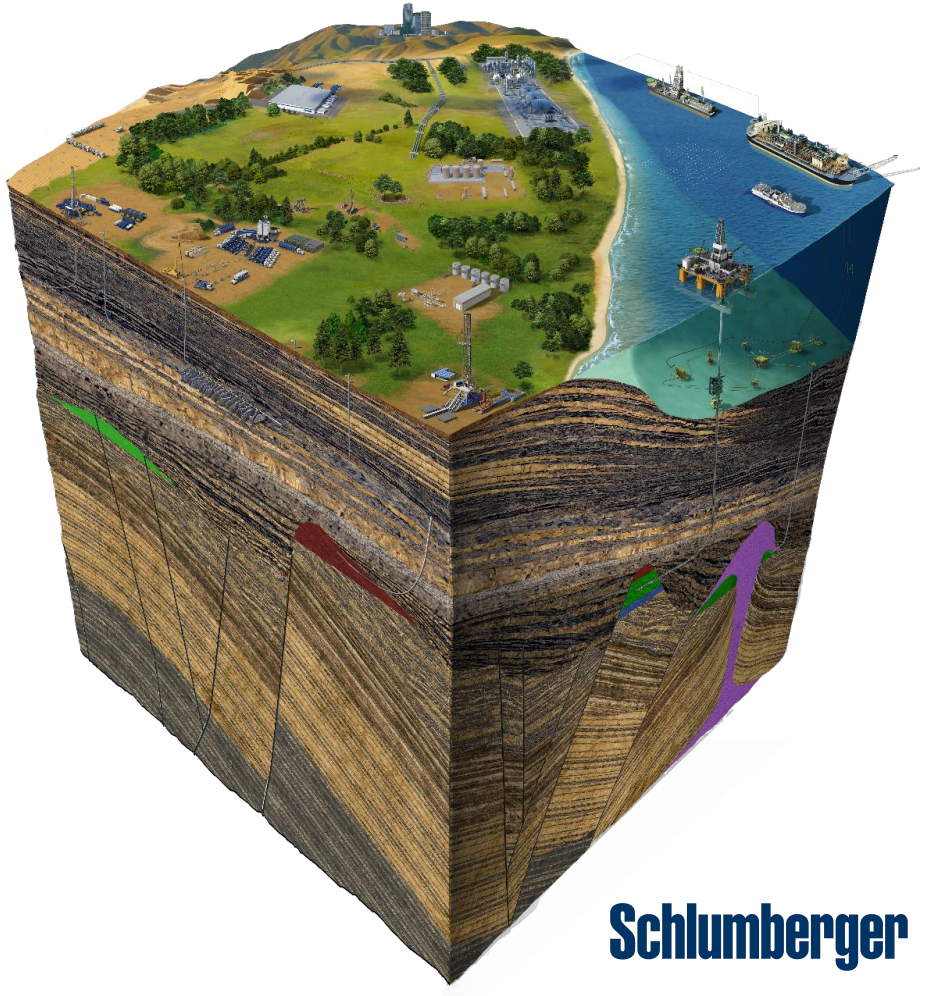


Tony Zhang
Senior Viz engineer
Schlumberger (prev)

Measuring the world

Since the early years of the 20th century, Schlumberger has been measuring the subsurface. Intellectual curiosity and commitment to research and technology are in our roots.

Today Schlumberger is the world's leading supplier of technology, project management, and information solutions to the oil and gas industry.



Google Cloud

Schlumberger

Data volumes in the oilfield

Types of Oilfield Data	Volume
Seismic Data – Onshore / Land Oilfields	6-10 TBytes / day x 6-12 month duration x 10s of projects/yr
Seismic Data – Offshore / Marine Oilfields	5-10 PBytes / project x 10s of projects/yr
Well Data – Drilling, Measurements, Testing, …	100s of GBytes / day x 1000s of wells/yr
Sub-surface Model Data	50-100 GBytes / project x 1000s of projects/yr
Production Data – Pumps, pipelines, networks, …	5 MBytes / day/ well x 10,000s of well/yr

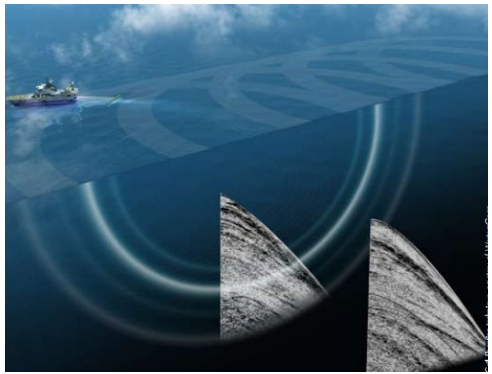
Google Cloud

Schlumberger

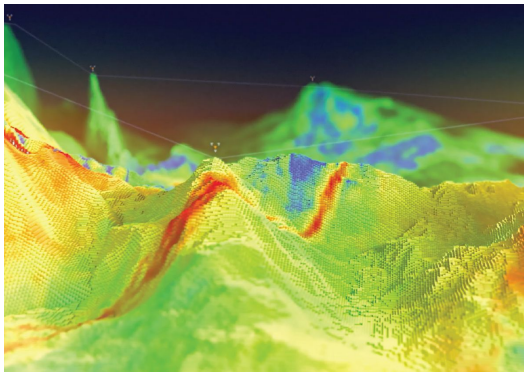
Seismic exploration

In a nutshell

Acquire data
indirect measurements about rocks illuminate subsurface



Build subsurface understanding
images, interpretations and geological models



- Large datasets recorded and generated
  - ➔ One job's input: 20-100TB
- Compute and IO intensive, concurrent
  - ➔ One small job uses hundreds of nodes for several day (100K core hours)
  - ➔ A large job uses thousands of nodes for 1-2 weeks (3-5M core hours)

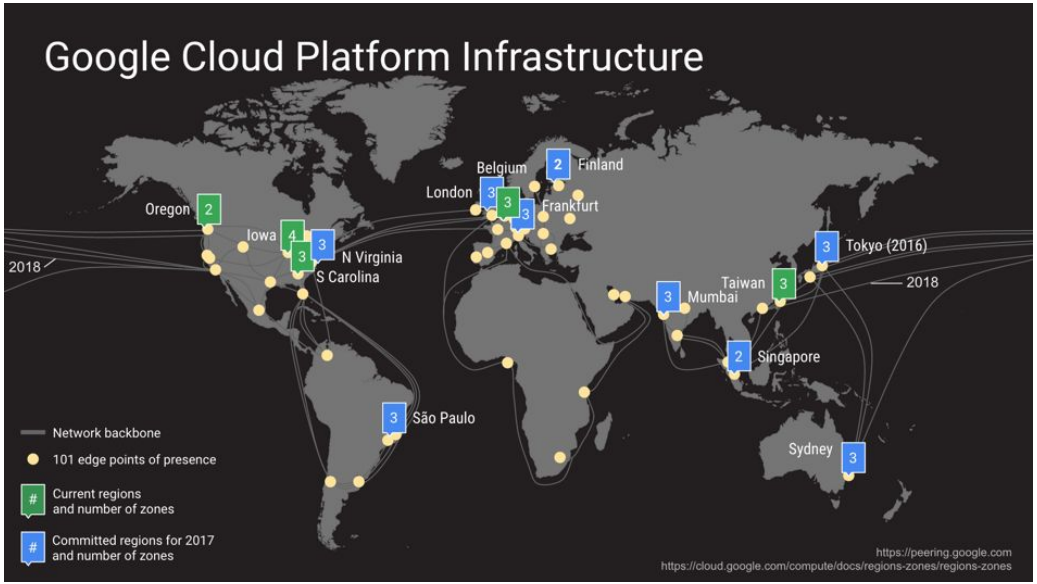
Google Cloud

Schlumberger

Simulation

Running in the Cloud

- Enables scalability by providing dynamic resources with short provisioning time
- Large volume of resources available
- Global presence
- Google technology as an enabler
- GPU nodes provide significant speedups
- Time to decision is a competitive advantage
- Increased number of scenarios
- More jobs in parallel at any given time (horizontal scale)



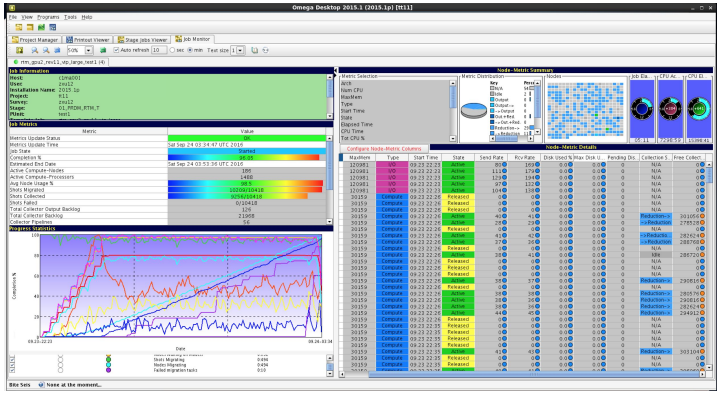
Google Cloud

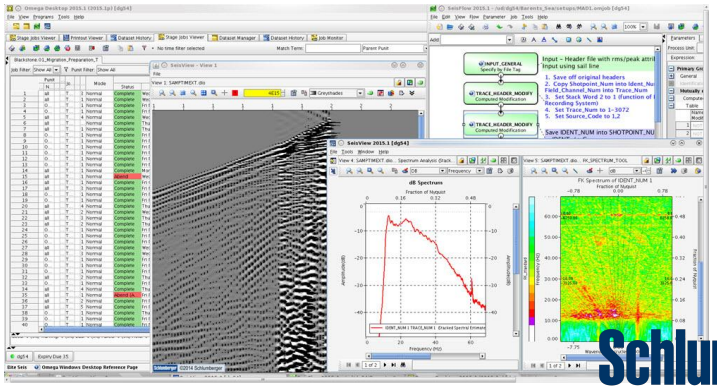
Schlumberger

Schlumberger's Seismic Software Stack

Tailored to the problem at hand

- **Scheduler and map-reduce like infrastructure**
  - Fault tolerance
  - Dynamic resource management
  - Avoiding MPI (due to fault tolerance requirements)
- **Custom solutions**
  - Job scheduler
  - Resource, data management layers
- **Schlumberger data centers around the world**
  - "Commodity" x86 clusters
  - GPU, infiniband for imaging applications





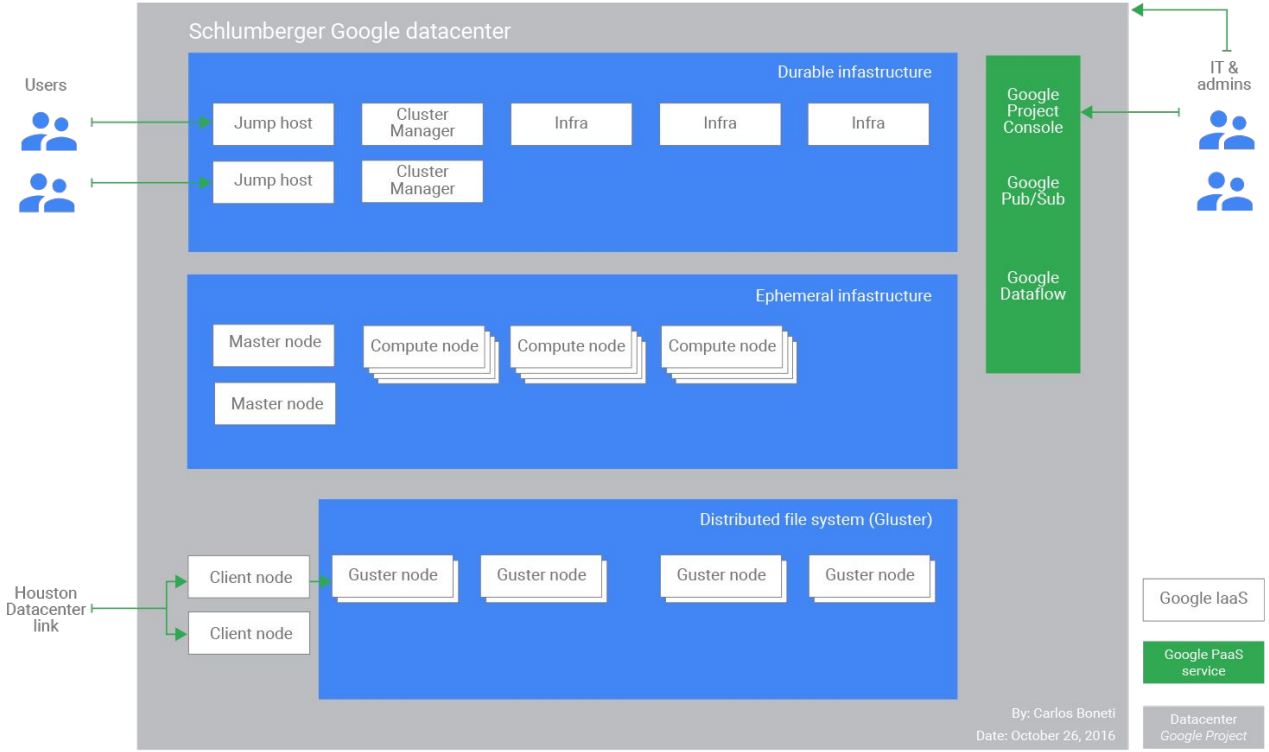
Google Cloud

Schlumberger

On the Google Cloud Platform

- **GPUs:** Provide great performance per $. Nodes with up to 8 GPUs minimize inter-node communication for processing
- **Computations:** instances dynamically created using deployment manager. VM sized per application for best combination of I/O bandwidth and CPU power
- **File systems:** persistent disks for scratch, gluster on GCE and GCS
- **Monitoring and control:** Monitoring done with internal tools and stackdriver. Control messages using pub/sub
- **Cloud Interconnect:** High speed link for data transfers before jobs
- **Secure:** Secure access, secure data, secure transmissions

On the Google Cloud Platform

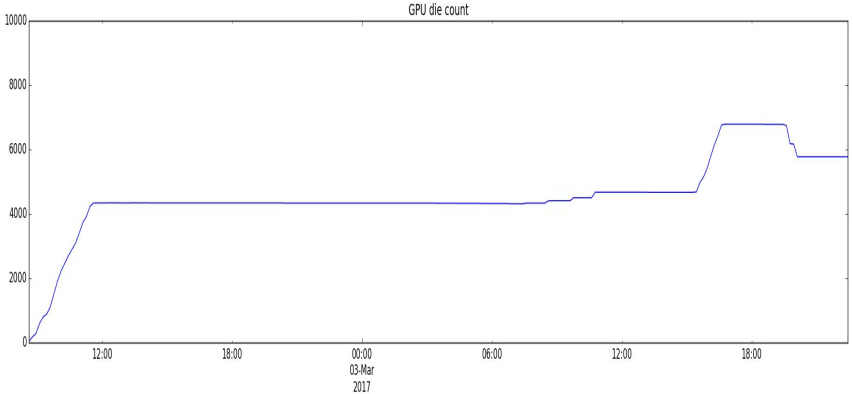


Google Cloud

Schlumberger

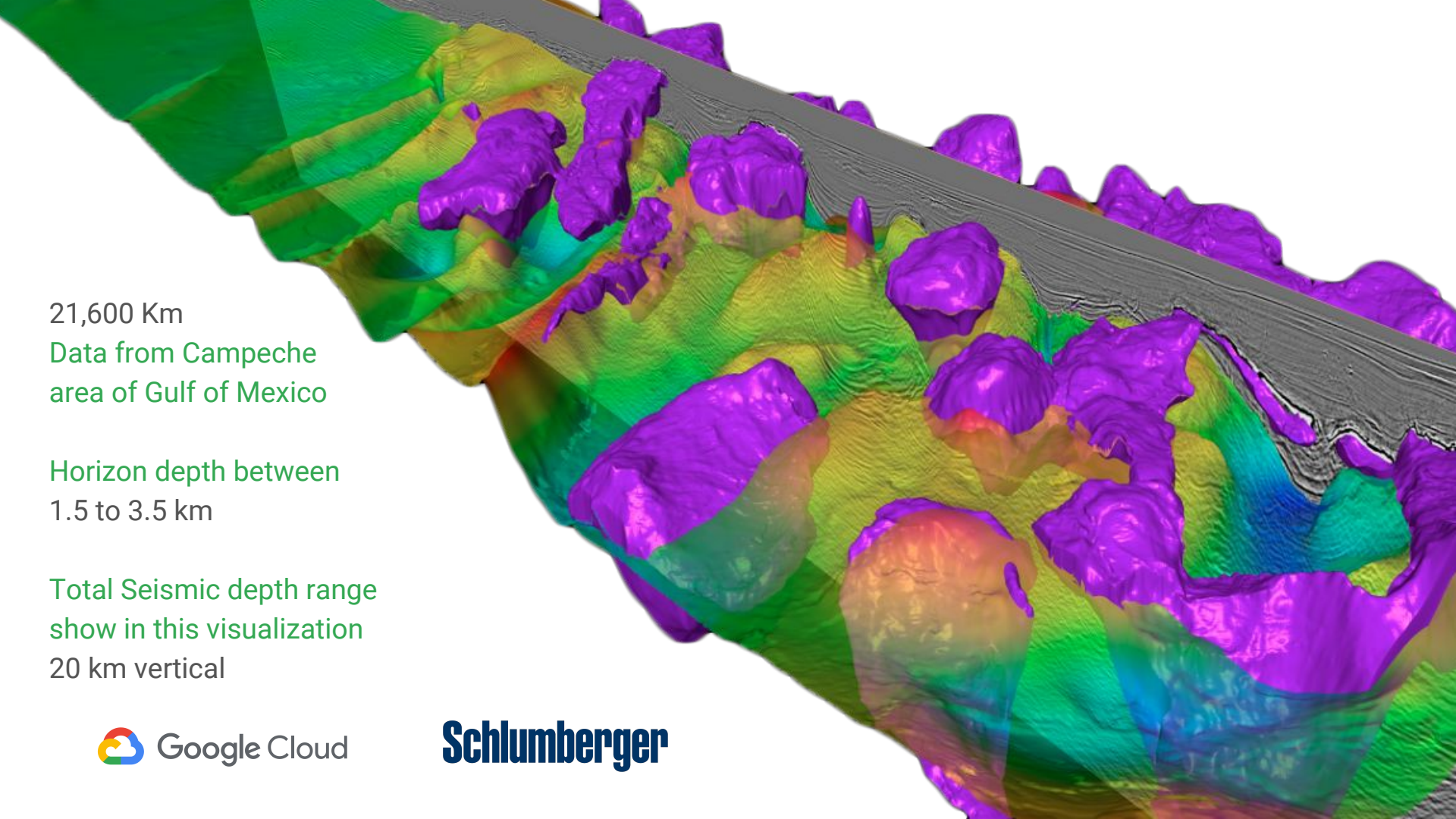
Example: A Seismic Imaging Job

- Reverse Time Migration (RTM) lasting for several days
- Variable number of nodes. Peak at 1750 nodes or ~7000 GPUs.
- Other jobs running in addition to this
- 2M+ CPU core hours on a single job
- 220GB/s on I/O



Google Cloud

Schlumberger



21,600 Km
Data from Campeche
area of Gulf of Mexico

Horizon depth between
1.5 to 3.5 km

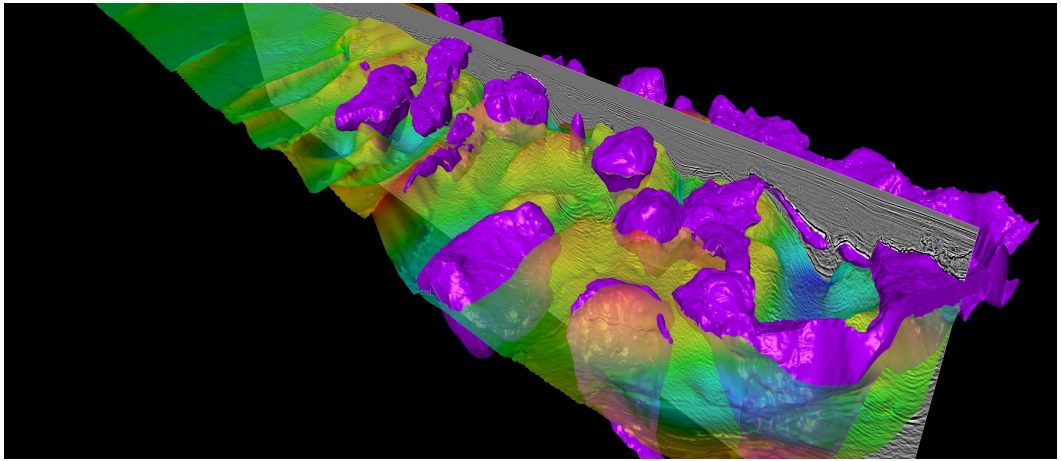
Total Seismic depth range
show in this visualization
20 km vertical

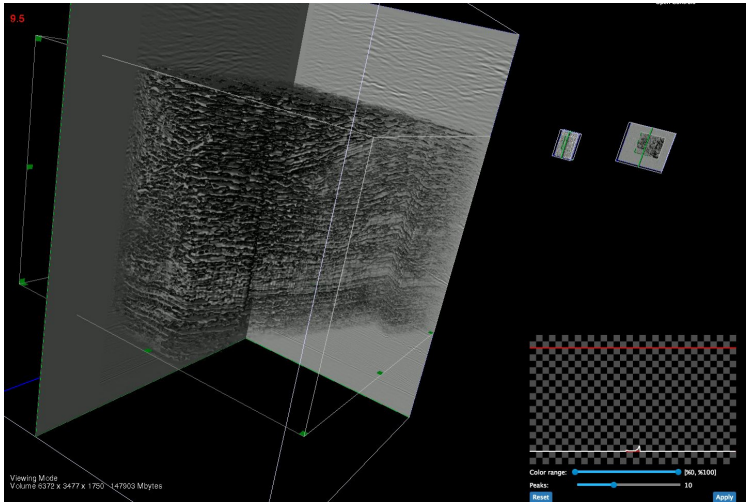
Google Cloud

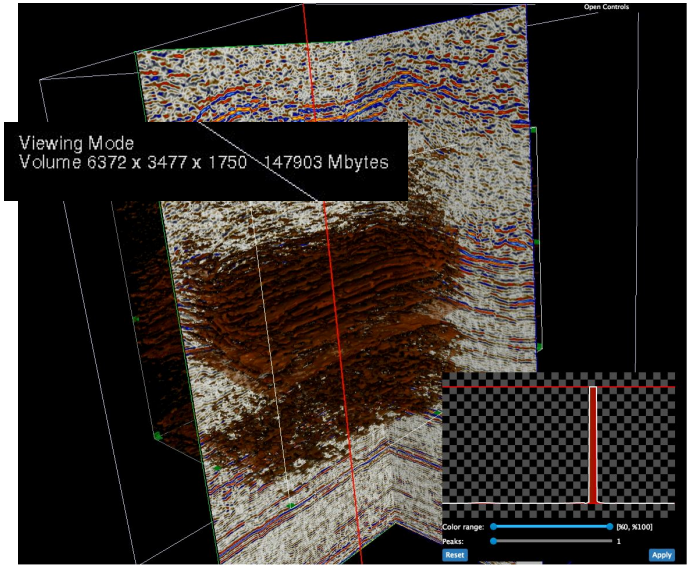
Schlumberger

Visualization

Visualizing the Data







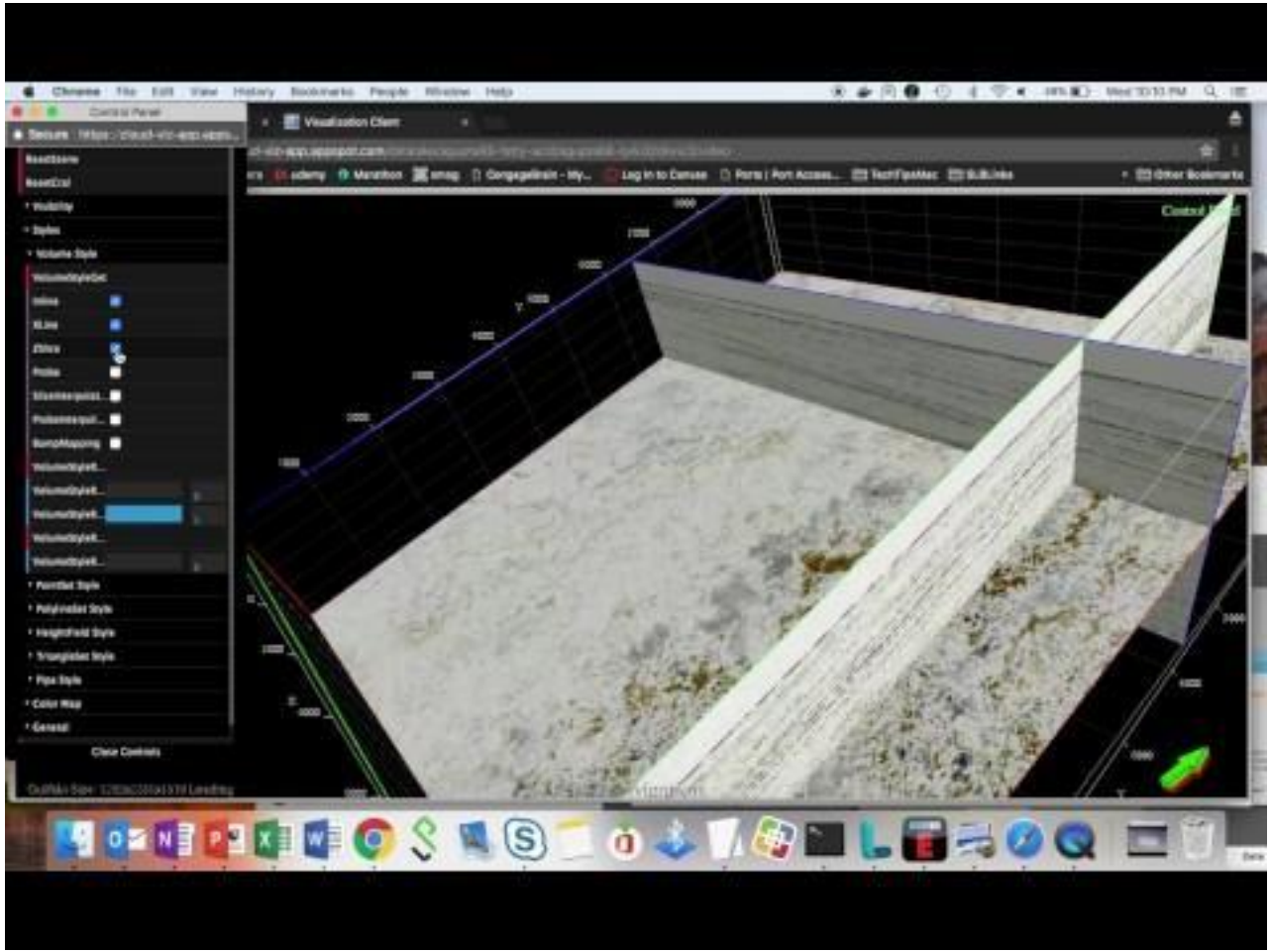
Schlumberger

Traditional

- **Durable infrastructure:** Very powerful machines that do computations and visualization of the data. Often difficult to delete and create in a fast way.
- **Local data or shared file system:** The time to copy the data from a central “cold” storage to a local copy can be measured in weeks (also due to non-technical reasons).
- **Planned collaboration and exploration of data:** The right data has to be brought to the right machine or file system encouraging isolated work and posterior presentation of results.

Viz API

- **Ephemeral VMs:** No data stored in the VMs makes provisioning and destroying VM fully automated and very fast. Because VMs only do rendering, they can be very lean. cluster of GPUs in Cloud on-demand
- **Data stored in GCS buckets:** No data is local to the VMs. Data is never copied locally and only the data needed for the visualization is streamed to the GPUs.
- **Spontaneous:** Data immediately available on any device, encouraging collaboration and exploration of analogous datasets across the earth. Instant access to data regardless of data volume
- User will sit anywhere, with variety of more or less powerful devices



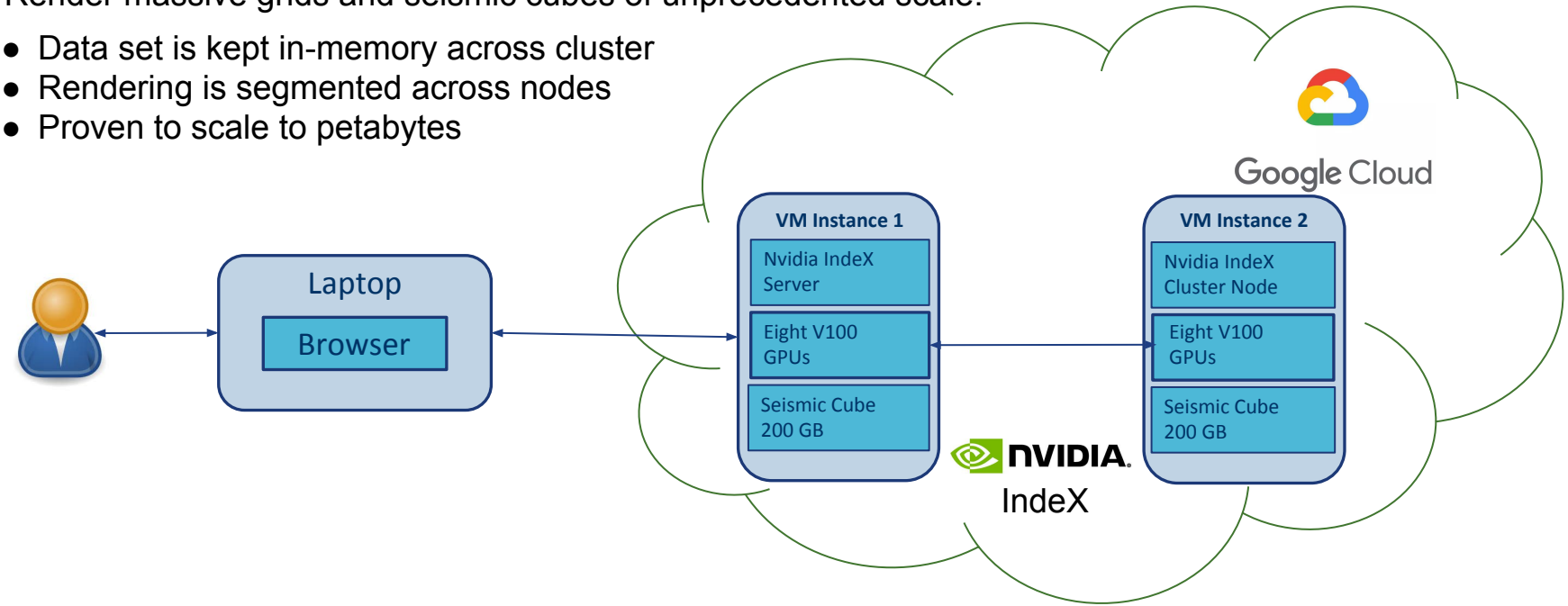
https://docs.google.com/drawings/d/18LxX68uvYAT8imN1i2rxdNXXouJEWLYhlq_vpY42lJg/edit

Schlumberger

Full resolution visualization with multiple GPUs in the cloud

Distributed rendering to visualize gigantic seismic cubes and grids.
Render massive grids and seismic cubes of unprecedented scale.

- Data set is kept in-memory across cluster
- Rendering is segmented across nodes
- Proven to scale to petabytes



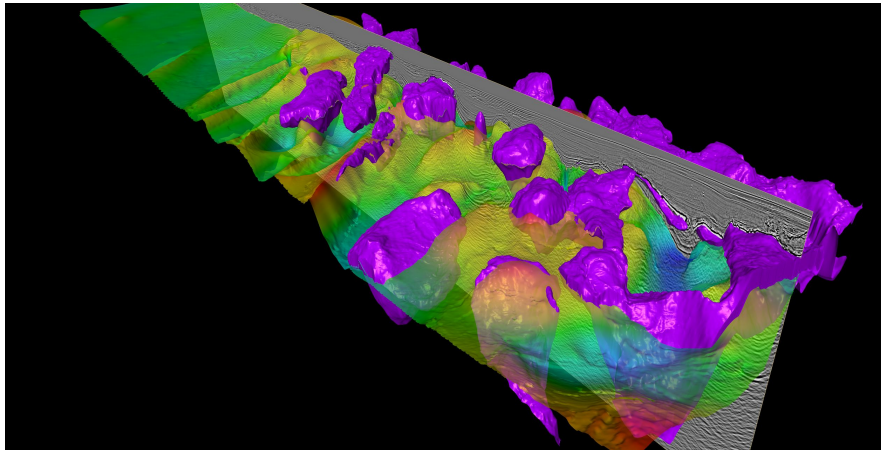
Schlumberger

In a nutshell

HPC in the Cloud

CPUs & GPUs for massive scale seismic processing in GCP

New workflows with cloud GPU native remote visualization



Schlumberger is able to lead innovation in oil and gas HPC thanks to rapid advancements in GPUs and quick adoption in the Google cloud

Google Cloud

NVIDIA

Schlumberger

Thank you.

https://cloud.google.com/hpc

Google Cloud