OCRE & Google (by Sparkle) Webinar

Jean-Pierre Aerts – Product Management

Agenda

Start	END	Topic	Who	Function
10.00	10.05	Belnet & OCRE Géant	Jean Pierre Aerts	Product Manager Belnet
10.05	10.10	Introduction Sparkle	Federico Italiano	Account Manager OCRE Framework
10:10	10:15	Introduction Google Team Belgium	Kurt Rommens	Leader Public Sector Belgium and Luxembourg - Google Cloud
10:15	11:30	OCRE Discount + Egress Solutions accessible through OCRE Use Cases Security, Data privacy	Sparkle & Google Team Focus on HPC by: Cristian Mezzanotte – Google Customer Engineer	
11.30	11:40	Q&A	ALL	

OCRE (Open Clouds for Research Environments) – General introduction

- Framework agreement (EU/NL procurement) for cloud services
  - The OCRE tender is conducted under the procurement law of the Netherlands (NL), called AW12. This procurement law conforms to EU procurement directive 2014/24/EU with some national implementation specifics.
- The framework agreements will be valid for four years until 2024.
  - Jan, 2021 – Dec 31, 2021 will have a 5 year term
  - Jan, 2022 – Dec 31, 2022 will have a 4 year term
  - Jan, 2023 – Dec 31, 2023 will have a 3 year term
  - Jan, 2024 – Dec 31, 2024 will have a 2 year term
- Can be immediately consumed without any commitment.

More information:

https://www.ocre-project.eu/national-research-and-education-networks-nrens-and-member-institutions

OCRE @ Belnet

- Belnet acts as a referrer and can guide you in making the right decisions and getting in contact with the right people.
- Contracts are signed between the customer (institute) and the partner/provider itself.
- Huge offer for IaaS, SaaS, PaaS. It also includes a “Earth Observation” tender for Copernicus satellite data processing services built on the DIAS platform.
- Want to find out more, please send an e-mail to: **cloud.geant@belnet.be**.

OCRE Providers & Partners

- OCRE: Open Cloud for Research Environments
  - Substantial discounts with 13 cloud providers
  - No Call for Tender needed, just sign a contract with the reseller
  - Regular work shops: subscribe to our OCRE FI@sh
  - Advice by Belnet: no strings attached, no costs attached
  - Cloud Connect: Express route for MS Azure

https://www.belnet.be/nl/diensten/cloud/geant-cloud-solutions

Belnet Event
Belgium
Google / Sparkle - OCRE 20 th
May 2021

Federico Italiano
Sparkle France, Paris

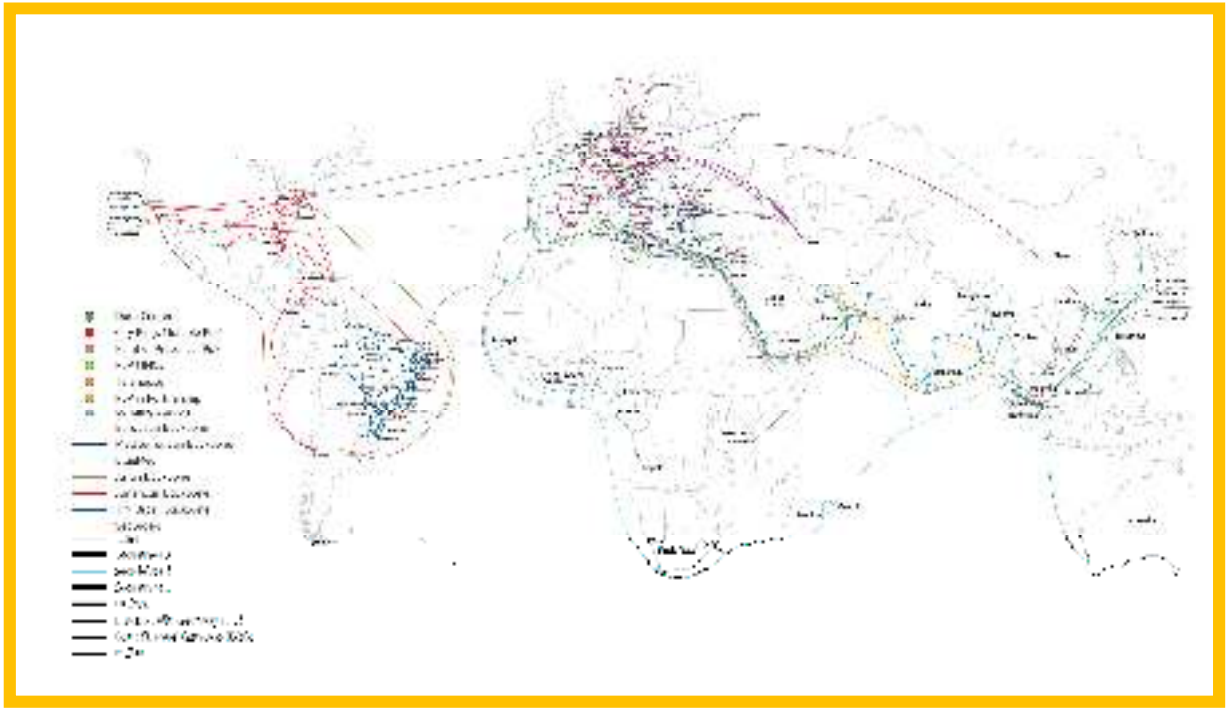
Kurt Rommens
Google Belgium, Luxembourg

Cristian Mezzanotte
Google EMEA, Zurich

SPARKLE

Google Cloud

Overview of Sparkle



GLOBAL CAPABILITIES

- **86** cities in 44 countries

- 159 PoPs
- >1,000 PoPs through partners

- 41,000 km/fibre PEB
- 58,500 km Americas Backbone
- 11,000 km Mediterranean Backbone
- 18,000 km/fibre Italian network
- more than 433,000 km submarine cables*
- ca. 53,000 km other fibre network*

- **8** Datacenters

- **800** Headcount Worldwide

Belnet

SPARKLE

Why Sparkle with Google

TIM

Enoovle

Telsy

Olivetti

SPARKLE

TLC Network - Italy

Google Cloud

Cybersecurity

Internet of Things

International Network

Google Cloud

TLC Network - Brezil

22 Datacenters

New DCs with Google

Google Cloud

Our presence in Belgium:
Avenue des Arts 6
Bruxelles 1210,
Belgium

Sparkle within the 2017 Géant Framework

Sparkle with Google in 27 Countries

Sparkle has been selected as unique reseller of Google Solutions in 27 Countries.

Albania	Germany	Netherlands
Armenia	Hungary	Poland
Austria	Ireland	Portugal
Belgium	Israel	Slovakia
Bulgaria	Italy	Slovenia
Croatia	Lithuania	Switzerland
Cyprus	Luxembourg	Turkey
France	Malta	Ukraine
Georgia	Montenegro	United Kingdom

Sparkle offer through OCRE

Google Workspace

Google Cloud

Professional Services

noovle

- POC
- Training
- Project
- Deployment

Applications

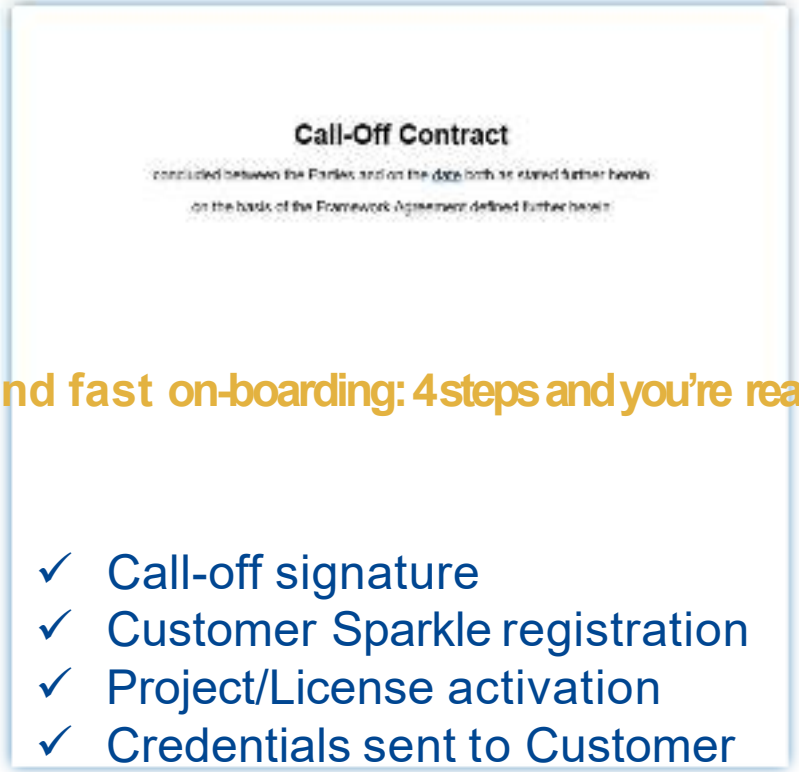
tellya

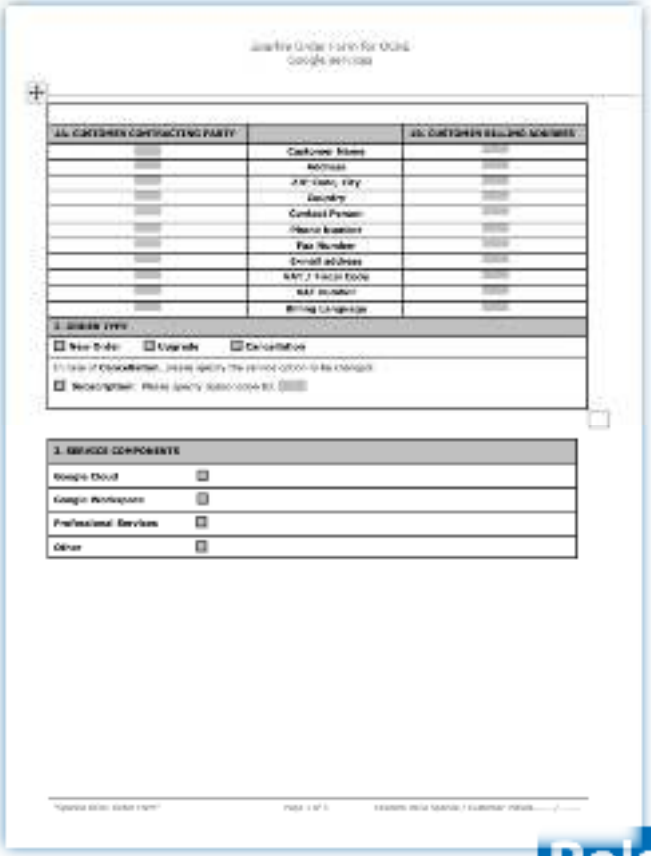
Quelix

OnTheMap

Belnet

How OCRE works - Procedure





Easy and fast on-boarding: 4 steps and you're ready to go.

- ✓ Call-off signature
- ✓ Customer Sparkle registration
- ✓ Project/License activation
- ✓ Credentials sent to Customer
- ✓ No minimum consumption

How OCRE works - Benefits

1. 6% Flat Discount **on the Public Price List** of the Google services.

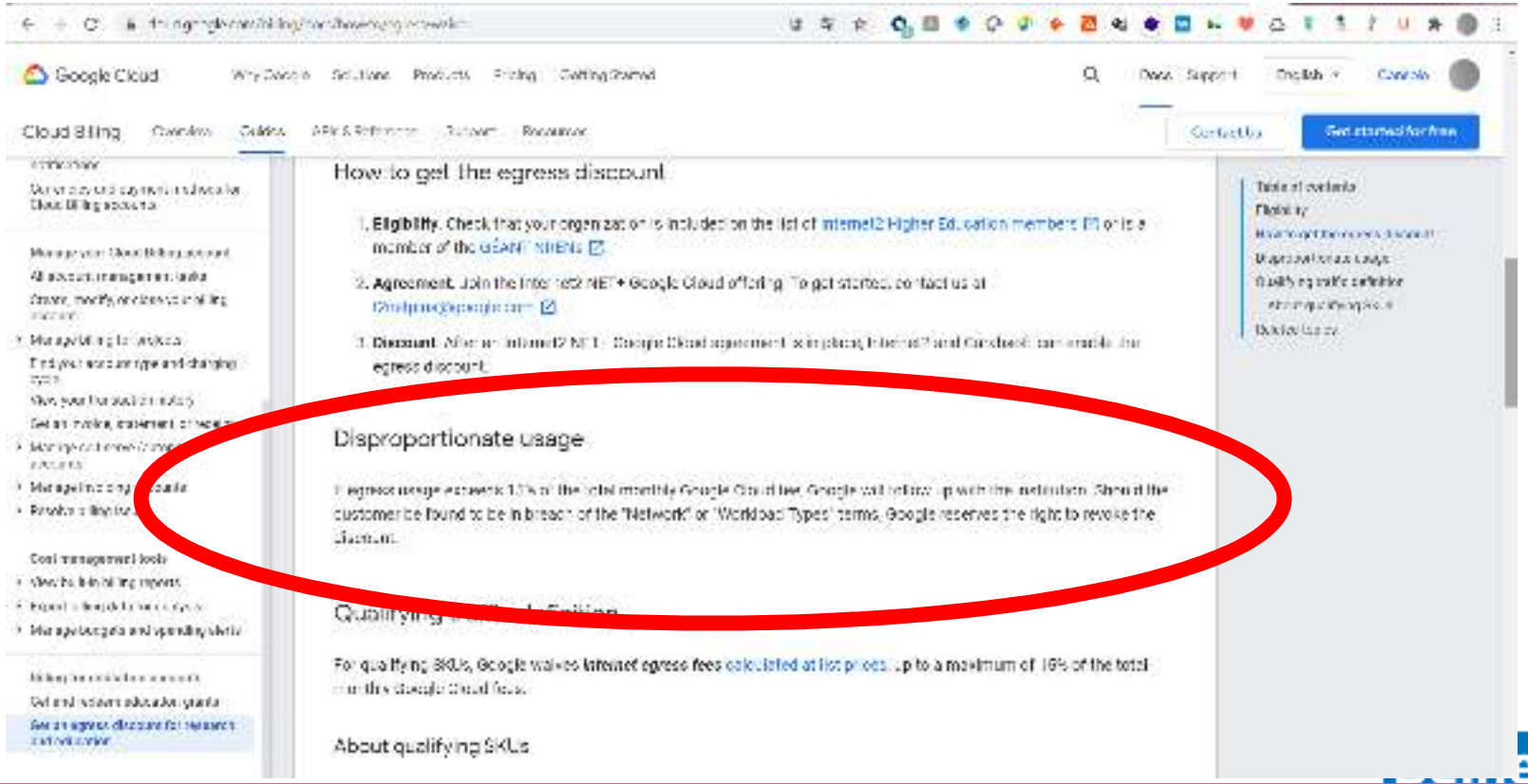
2. Volume and term based discount

Sparkle propose a further discount based on revenues volume and contract term as per the below scheme:

Volume and term based discount				
Volume discounts	Tiers/Term	1Y	2Y	3Y
Total contract value	> 350k€	1%	1,50%	2%
Total contract value	> 700K€	1,50%	2%	2,50%
Total contract value	> 850K€	2%	2,50%	3%

3. Full Data Egress Waiver: Sparkle applies no charge to traffic generated from Google resources (with no limit on the contract value).
https://cloud.google.com/billing/docs/how-to/egress-waiver?hl=it

Focus on Data Egress

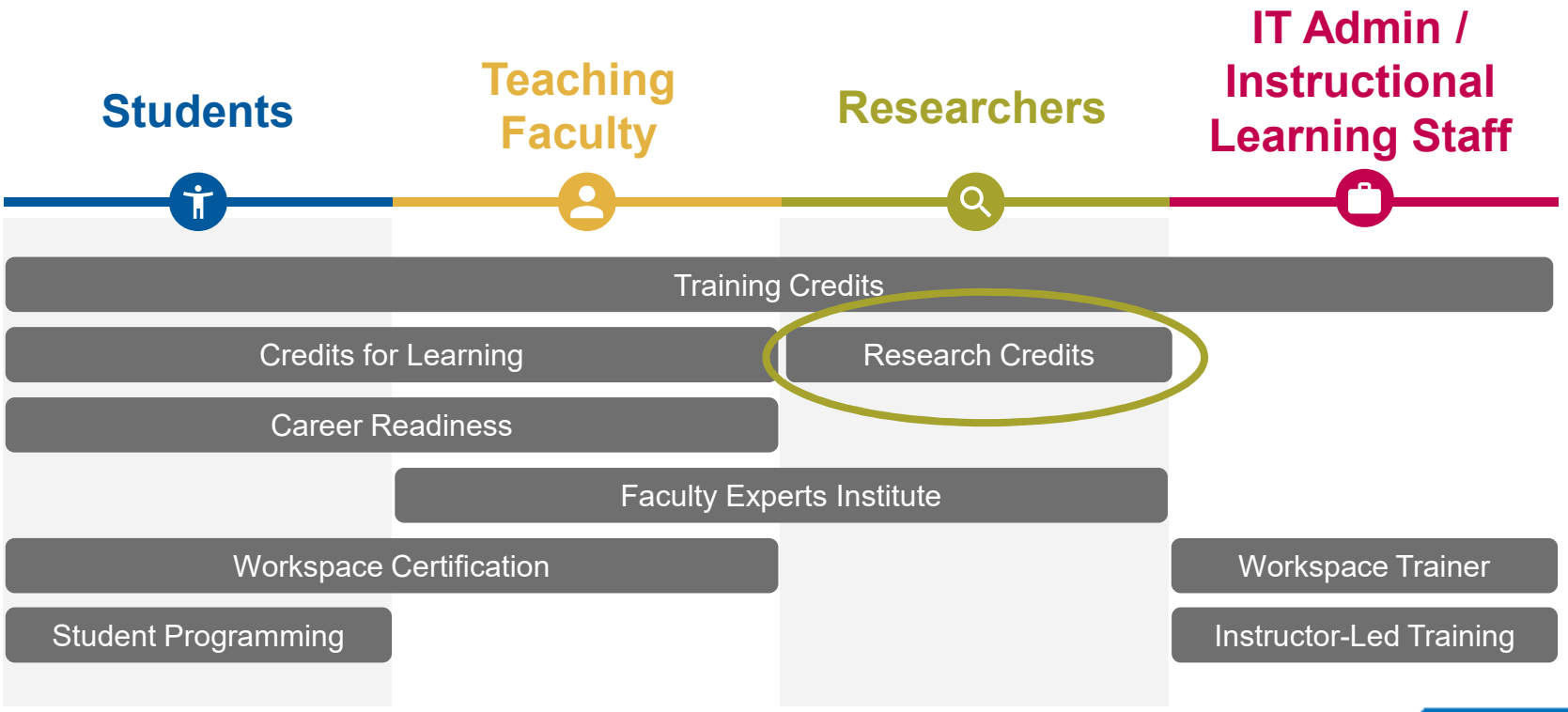


Focus on Data Egress

Role	Faculty, administrators, and academic researchers using Google Cloud
Workload Types	Allowed Research or academic workloads Hosting departmental websites NOT Allowed Massive online open courseware (MOOC) 1.Live-streaming/ Media streaming 2.Hosting websites for commercial purposes outside of research & teaching.

Higher Education Programmes

Google Cloud



Belnet

Sparkle offer through OCRE

Google Workspace

Google Cloud

Professional Services

noovle

- POC
- Training
- Project
- Deployment

Applications

tellya

Quelix

OnTheMap

Belnet

Google Workspace – Formerly GSuite

Belnet

SPARKLE

Focus on On-Site Professional Services

- ✓ **Training** on cloud computing methodology and best practices within Google technology;
- ✓ **Analysis** of the organization's goals over the next 1-3 years;
- ✓ **Assessment** about the critical applications for achieving the identified goals, and the IT infrastructure for legacy applications: hardware, software, security and cloud-based services currently in use;
- ✓ **Design** of the Cloud Architecture for the best fitting of the requirements;
- ✓ **Support to the Migration** in the Cloud of the target applications;
- ✓ **Management** of the Cloud infrastructure.

Belnet

Quelix for University of Naples - Private Search Engine

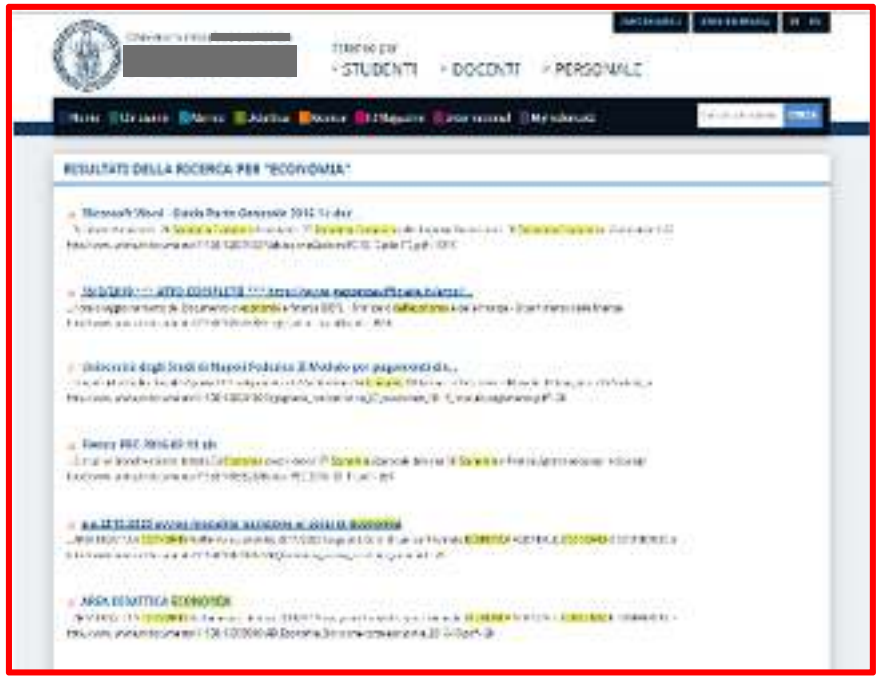
Solution

Implementation of a single search engine providing all vertical portals per faculty.

Implementation of a Onebox service for searching contacts in the Address Book

Products used

- Quelix with ON-GCP installation
- Database connector



tellya
CONVERSATIONAL INTELLIGENCE

AI 24/7 Virtual Assistant value

How this will help your teams

+90% common questions answered automatically

Personalized answer based on student profile, +120 languages Any-time, Multi-channel

24/7 Virtual Assistant

Transfers case along relevant information about the student, the request, background of conversation and recommended solutions

Critical or unique issues are transferred to a real advisor

Academic Advisor or Support

Actionable Insights

- Key issues on campus
- Suggestions from students (opportunities for improvement)
- Analysis based on type of student, college, program, etc
- Customizable dashboards
- Trends and Analytics

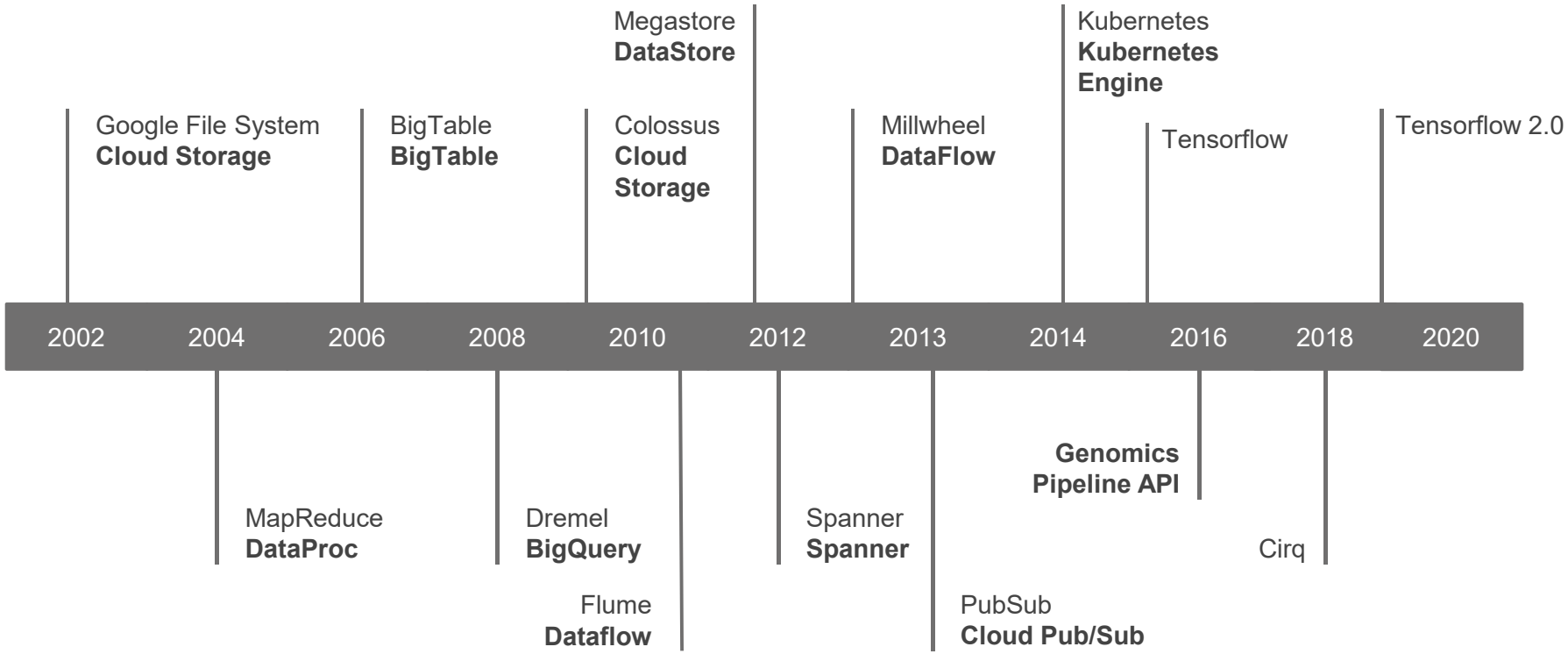
Digitization of the Campus

High Performance Computing on GCP

Cristian Mezzanotte
Customer Engineer, Google Cloud

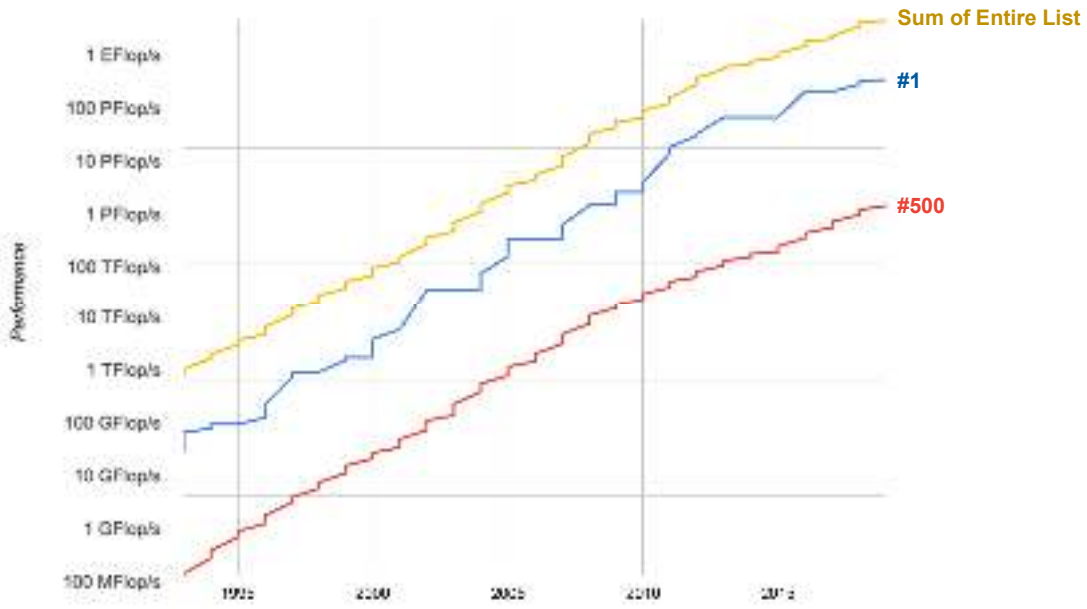
Google advances scientific computing

OSS / Whitepaper
Cloud Solution
Partner Solution



Google Cloud

Compute needs are growing



Top500.org Compute Growth

Data is now BIGGER than big

2008		2018
Los Alamos National Laboratory - RoadRunner		Oak Ridge National Laboratory - Summit
3PB Capacity - Panasas Storage	**83x**	250PB Capacity - IBM Spectrum Scale
55 GB/s throughput	**45x**	2.5 TB/s throughput

What is High Performance Computing?

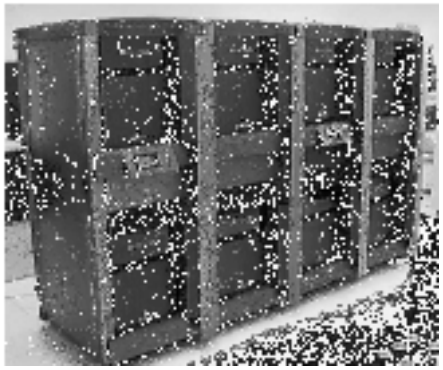
HPC is a broad market term for engineering, scientific and forecasting workloads that require parallel computing processing

Parallel computing helps speeding up the processing of challenging problems which would require too long in traditional computing infrastructures

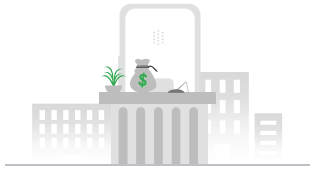
Traditionally, parallel computing is enabled by Supercomputers, which are special systems designed to maximize computation speed and performance

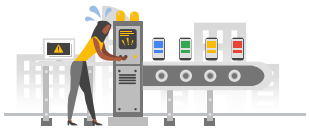
https://en.wikipedia.org/wiki/Supercomputer

HPC evolution



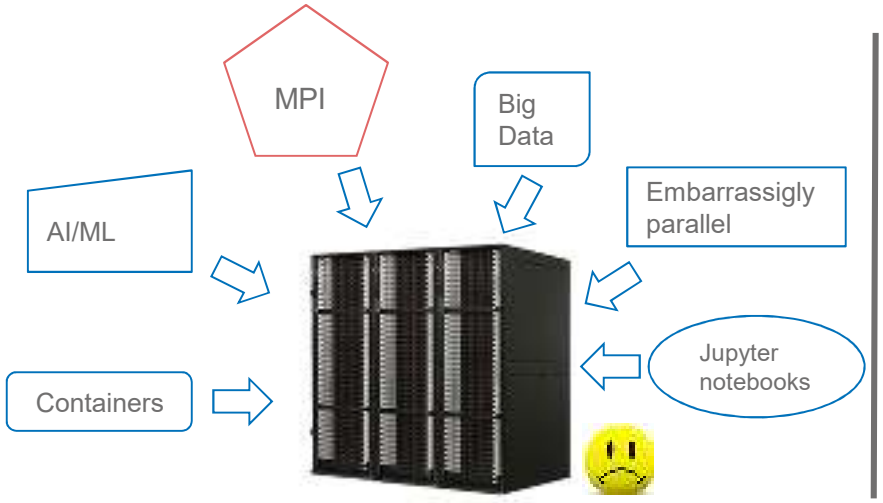
HPC drives R&D and innovation across industries





Belnet
dedicated connectivity

The Google Cloud Difference



On-Prem model: Higher wait times, downtimes, one size fits all aging hardware

Match resources to requirements

GCP model: No wait times, no downtimes, unique hardware and services tailored to researcher needs

Democratize **h**igh **p**erformance **c**omputing
and make it universally accessible and useful.

Urgent HPC

Most researchers do not have access to the HPC resources needed to study complex systems

Traffic Management is a major hurdle to efficient emergency evacuations

Models of evacuation traffic patterns during hurricanes are valuable for evacuation planning

Reducing traffic congestion saves lives, time, money, and the environment

DDN STORAGE + Google Cloud

Top 2 Cloud-Based Lustre Systems on the IO-500

Build your own world-class Lustre filesystem with DDN EXAScaler in the GCP Marketplace.

38	PGS	Lustre on GCP	Google	Lustre

Just search “**Lustre**”.

One customer achieved **#38**! Can you do better?

IO500

	information				
	list id	institution	system	storage vendor	filesystem type
1	sc19	WekaIO	WekaIO on AWS	WekaIO	WekaIO Matrix
2	sc19	Intel	Wolf	Intel	DAOS
3	sc19	National Supercomputing Center in Changsha	Tianhe-2E	National University of Defense Technology	Lustre
4	sc19	NVIDIA	DGX-2H SuperPOD	DDN	Lustre
5	sc19	University of Cambridge	Data Accelerator	Dell EMC	Lustre
6	sc19	CEA	Tera-1000	DDN	Lustre
7	sc19	CSIRO	bracewell scratch2	Dell/ThinkParQ	beegfs
8	sc19	State Key Laboratory of High-end Server & Storage Technology (HSST)	[illegible]	INSPUR	BeeGFS
9	sc19	Google Cloud	EXAS-GCP-PD-SSD	Google Cloud	Lustre
10	sc19	DDN	AI400	DDN	Lustre
11	sc19	Google Cloud	EXAS-GCP-PD-STD	Google Cloud	Lustre

IO500 2019-11 SC19 List
https://www.vi4io.org/io500/start

Belnet dedicated connectivity

The Google cloud Difference

Commitment to an open cloud

Google believes in the power of open-source to give customers **the freedom to choose** where to deploy their workloads and **avoid vendor lock-in.**

- Researchers and admins are familiar with open source - easy adoption of GCP
- Next-Generation hybrid and multi-cloud solution

Secure by design

One of the largest **privately managed** networks with over 100 points of presence across the globe. **Zero-Trust design.**

- Built-in HIPAA/FedRAMP/PCI compliance across all zones.
- Auditability and Access Transparency logs

Optimize spend

Unique flexibility to provide the **most cost-effective platform.**

- Automatic discounts and right-sizing deliver **35% savings on average**
- **Custom Machine Types, Preemptible VMs (fixed 80% off), Preemptible GPUs/TPUs (fixed 70% off), Compute Optimized VMs** to meet your needs.

The Google cloud Difference

Access to the latest technology

Harness the latest hardware to improve efficiency as soon as it becomes available. GCP was first to cloud with the **latest Intel Cascade Lake, AMD EPYC Rome CPUs and NVIDIA A100 GPUs.** GCP is the only provider with **TPUs** (Up to 100+ petaflops per pod)**.**

- On-demand elastic access to the most powerful hardware for each workload.

Innovate faster

Scale HPC workloads to previously unattainable dimensions. **On demand elastic resources** that fits perfectly to your workload.

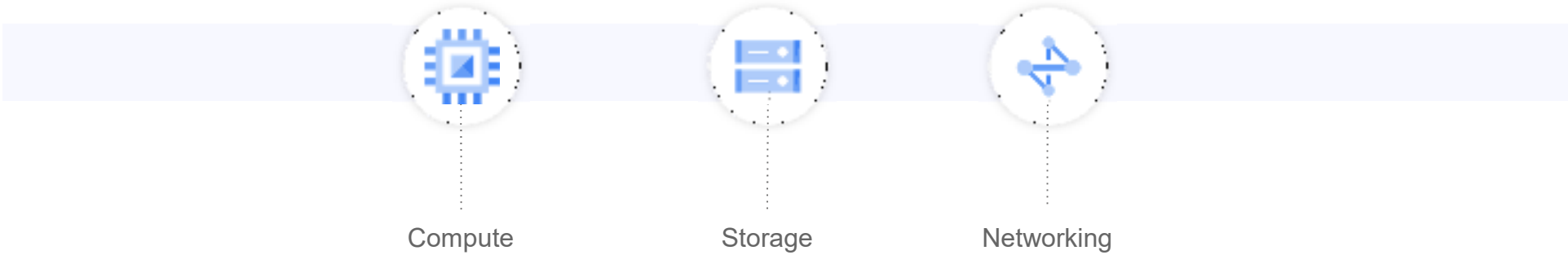
- On-demand bursting to cloud to answer bigger questions faster.

The best of Google

Google Cloud brings **the best of Google's innovation to transform operations and user experiences**, and to accelerate innovation and discovery.

- Embedded AI/ML to unlock new research areas and Win more grants.
- Commitment to carbon neutrality and sustainability

HPC Infrastructure



Google Cloud

Belnet
dedicated connectivity

Computing

Google Cloud

Belnet
dedicated connectivity

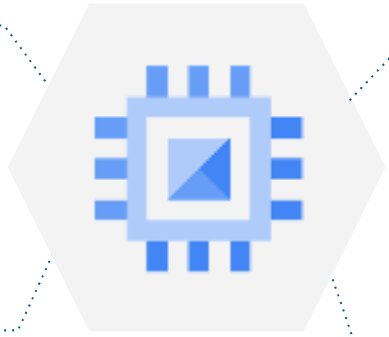
Google Compute Engine

General Purpose Instances

- N1, N2, E2 VM Families
- Max per instance:
  - 416 vCPUs
  - 12TB RAM
  - 9TB Local SSD
  - 8 GPUs

HPC Instances

- C2 VM Family
  - 60 vCPU (Intel Cascade Lake), 3.8 GHz Turbo, 240GB RAM
- N2D VM Family
  - 224 vCPU (AMD EPYC Rome), 2.7GHz Turbo, 896GB RAM
- vNUMA, Overhead Reduction



GPU Instances

- A2 VM Family
  - 96 vCPU (Intel Cascade Lake), 3.8GHz Turbo, 1,360GB RAM
  - 16 NVIDIA A100 GPUs (40GB)
  - vNUMA, Overhead Reduction
- N1 VM Family
  - 96 vCPU (Intel Skylake), 2.0 GHz, 624GB RAM
  - Max 8 GPUs per instance
  - NVIDIA K80, P100, P4, T4, V100

Cost Efficiency

- Preemptible VMs
- Custom Machine Types
- Auto-Scaling HPC Clusters
- Rightsizing Recommendations

Belnet
dedicated connectivity

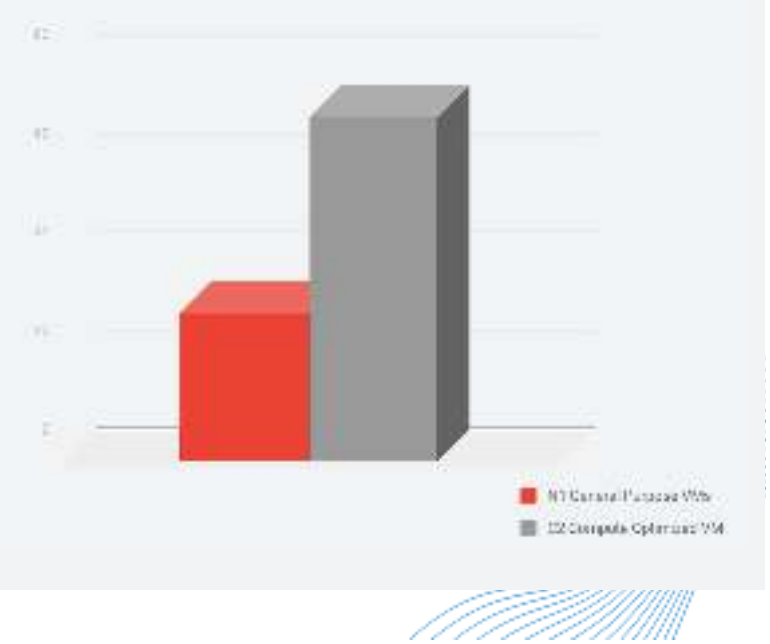
Product Specs

- High frequency 3.8GHz all-core-turbo
- Up to 3TB Local SSD
- Up to 32 Gbps Network Bandwidth

Highlights

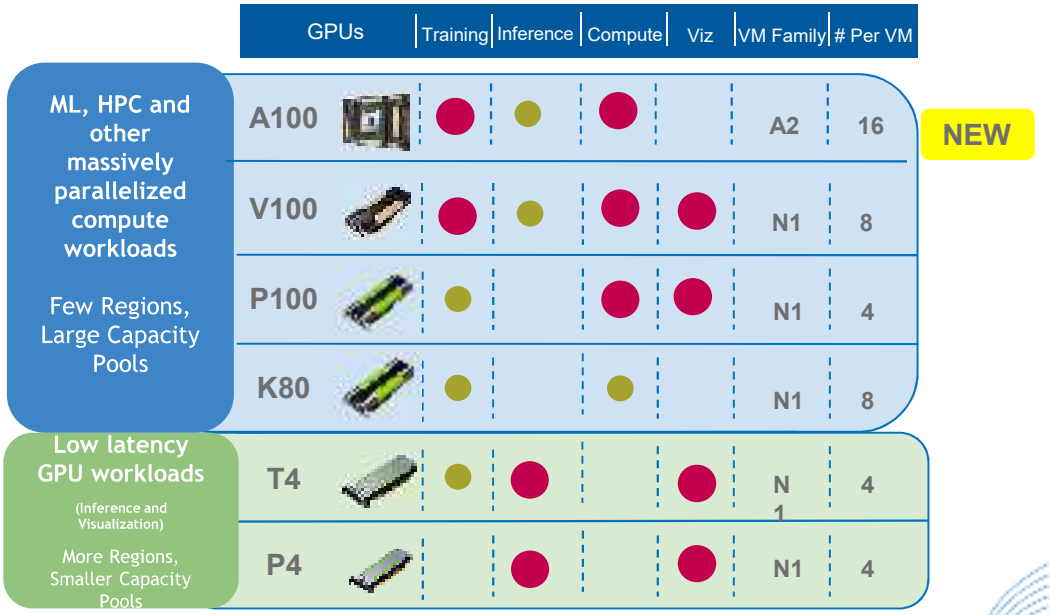
- NUMA-aware for Performance (vNUMA)
- Higher Performance per thread
- Isolation for latency sensitive workloads
- ***New!*** Placement policies for rack level colocation to reduce inter node latency in your HPC cluster

40% Higher Performance Per Core



Google Compute Engine - GPUs

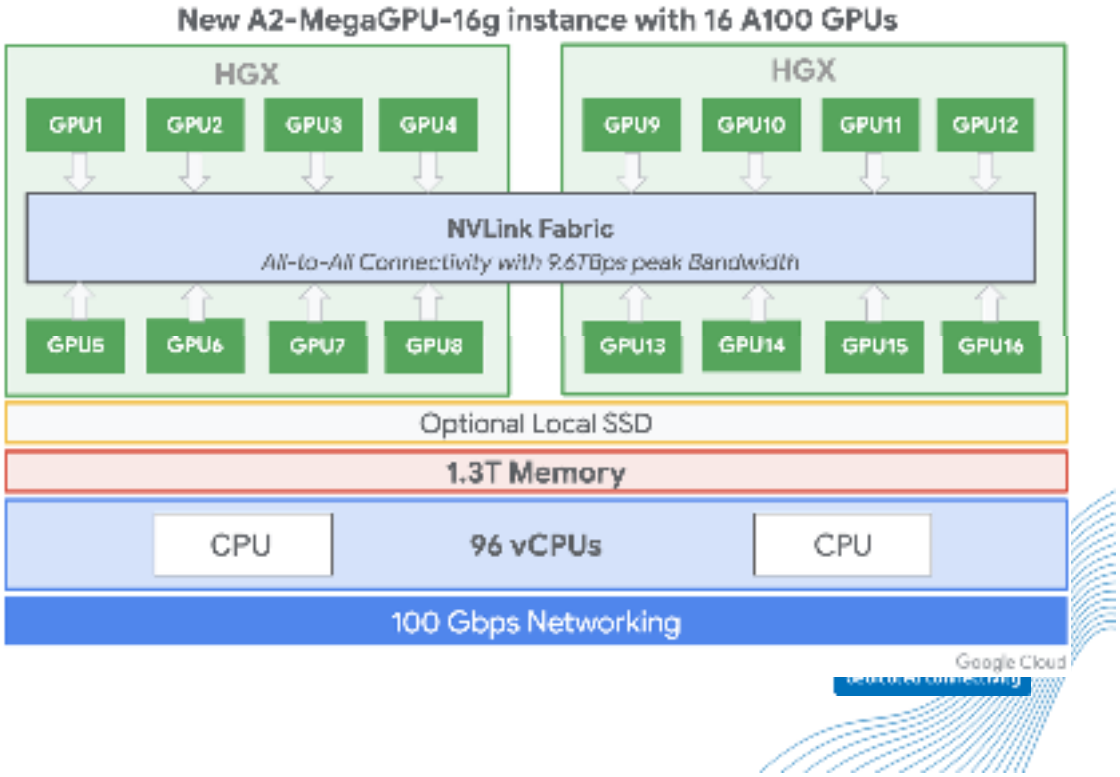
- Attached directly to the VM via PCIe x16, with NVLink and NVSwitch to achieve the best possible performance
- Per-second billing, Preemptible support (~70% off)
- First to market with NVIDIA A100, T4 GPUS



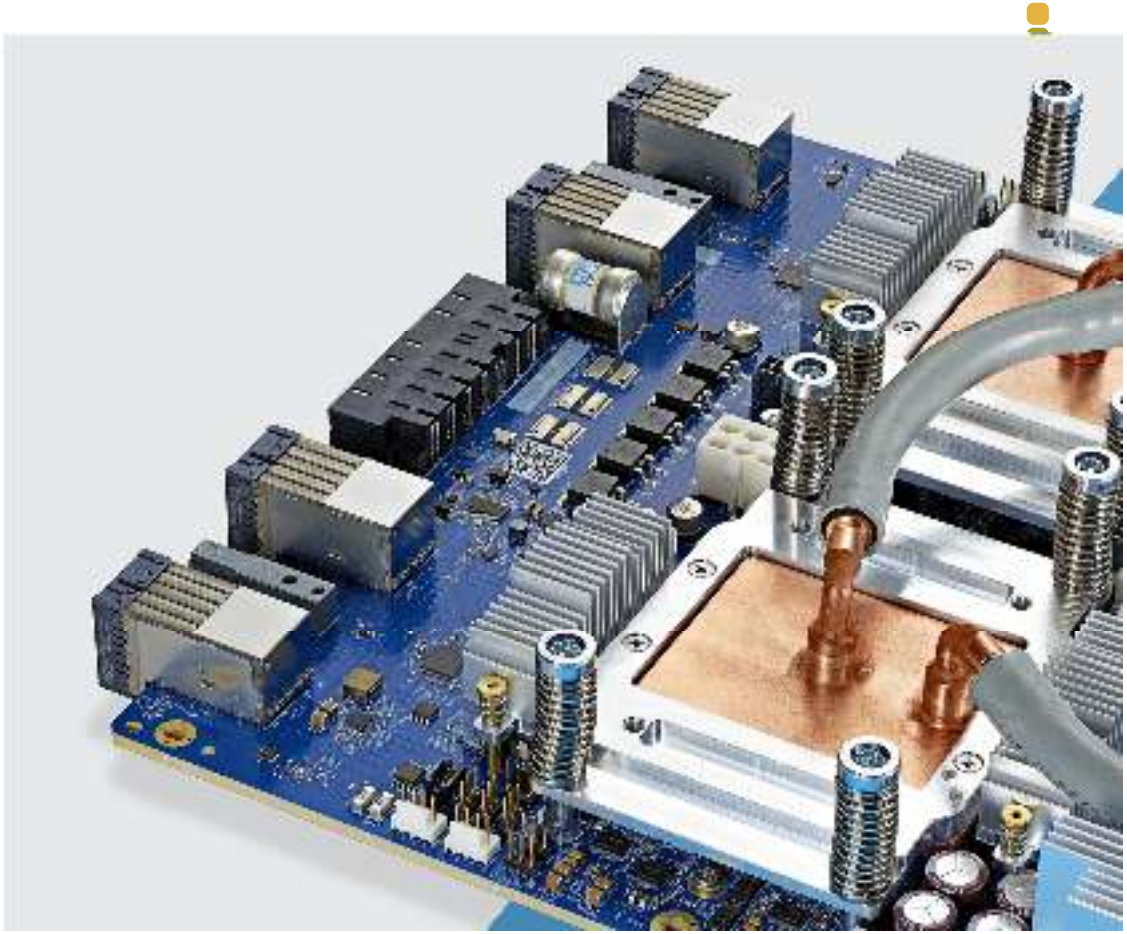
GCP A2 VMs INCLUDE UP TO 16 GPUs

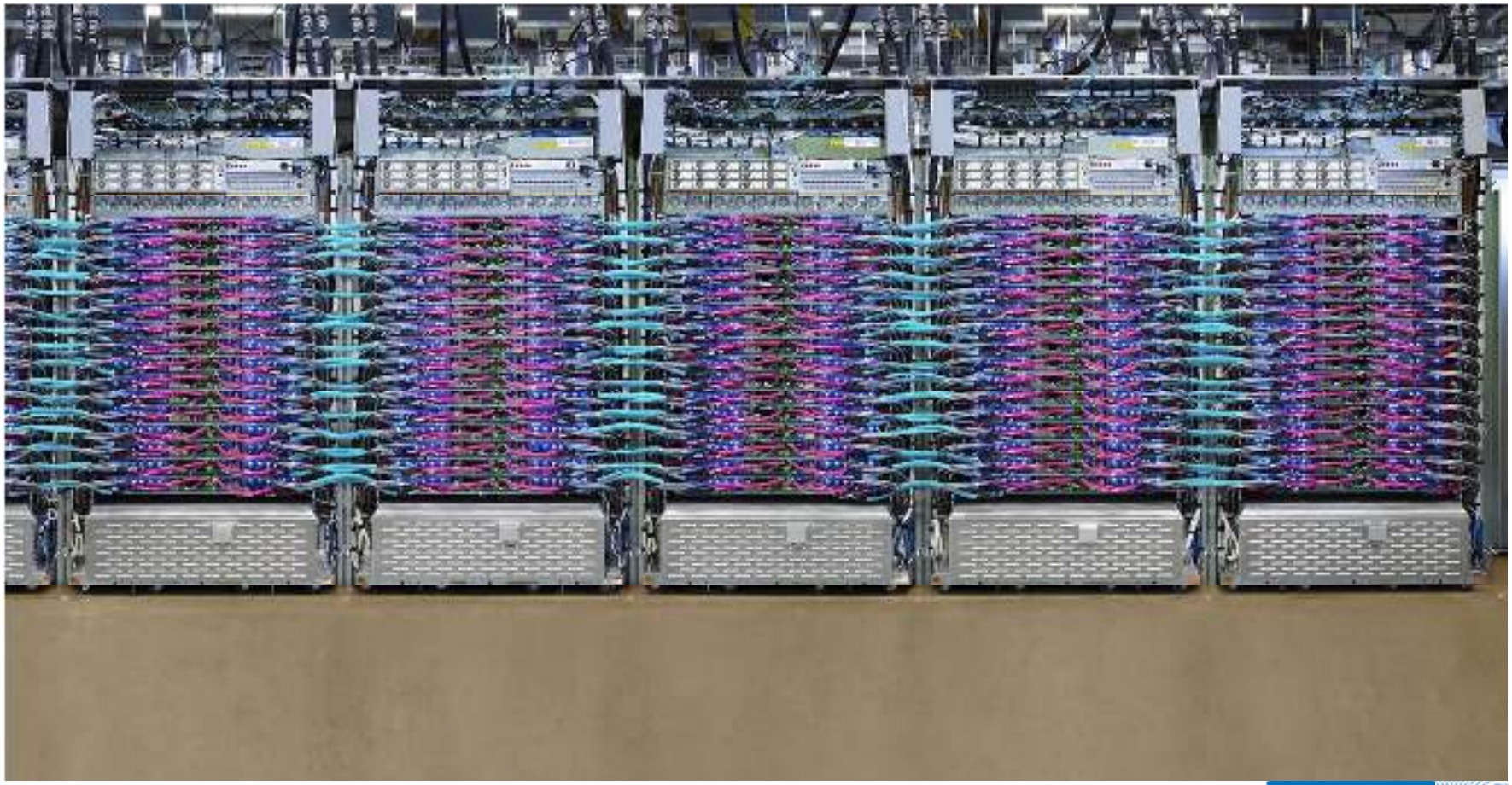
Introducing the *A2-MegaGPU-16g* VM

- 2 HGX Boards
- 2.4TB/s total NVLINK BW w/ 16 A100 GPUs
- *ONLY* CSP offering up to 16 GPUs in a single node
- Effective performance up to 10 petaflops FP16 or 20 PetaOps of int8 in a single VM *(using the new sparsity feature)*



Cloud TPU v3





Cloud TPU v3 Pod: >100 PFLOPS

Storage

Google Cloud

Belnet
dedicated connectivity

Which storage type?

In Memory	Relational		NoSQL		Analytical					
										
Memorystore	Cloud SQL	Cloud Spanner	Firestore	Cloud Bigtable	BigQuery	Cloud Storage	Persistent Disk	Local SSD	Filestore & Partners	
Managed Redis & Memcached	Managed MySQL and PostgreSQL, and SQL Server	Scalable relational database	Serverless, scalable, document store	Low-latency, scalable key-value and wide-column store	Enterprise DW					

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, Dell EMC, DDN, Lustre, and more
Move petabytes to GCS with the Data Transfer Appliance

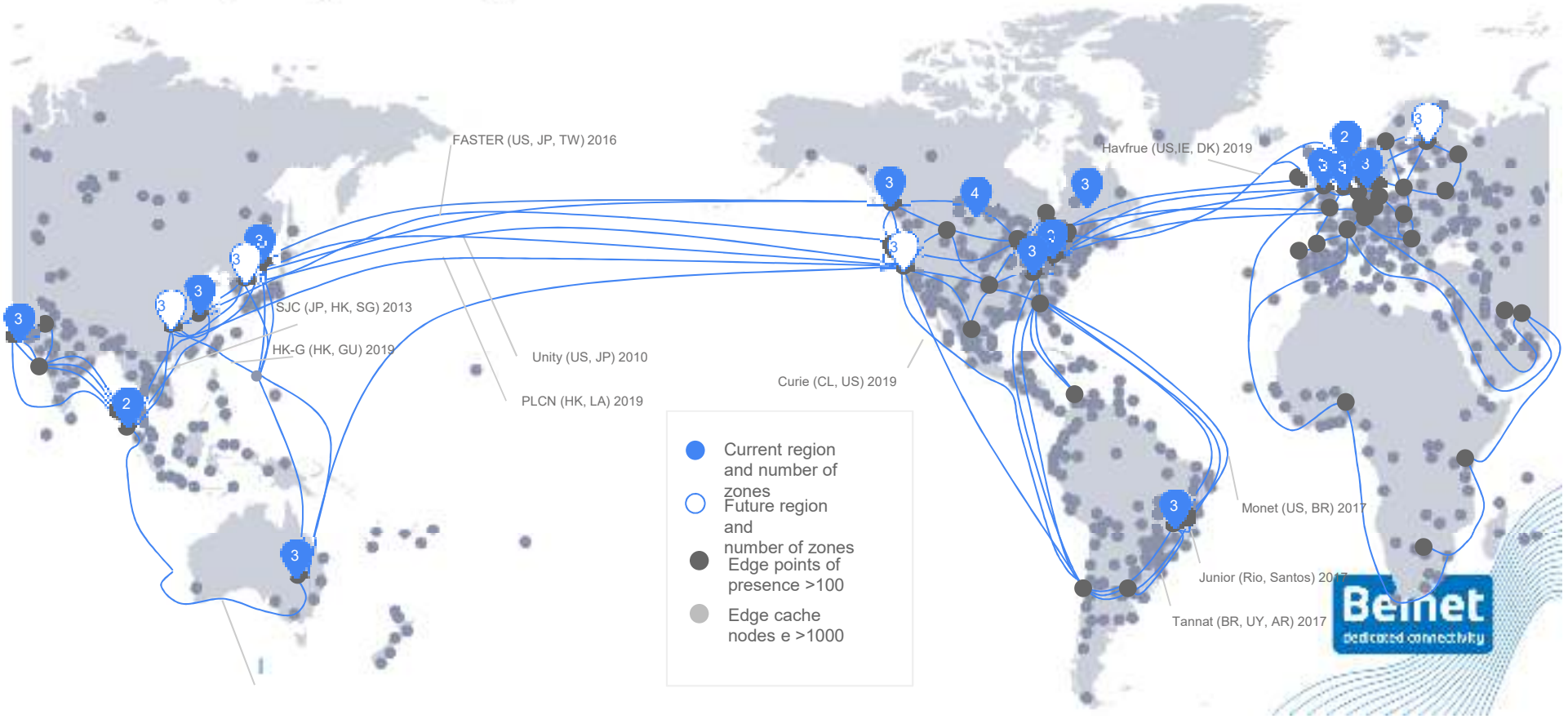
Networking

Google Cloud

Belnet
dedicated connectivity

Global network infrastructure

The largest cloud network: 100,000s of miles of fiber optic cable, 8 subsea cables
More edge and peering points than any public cloud



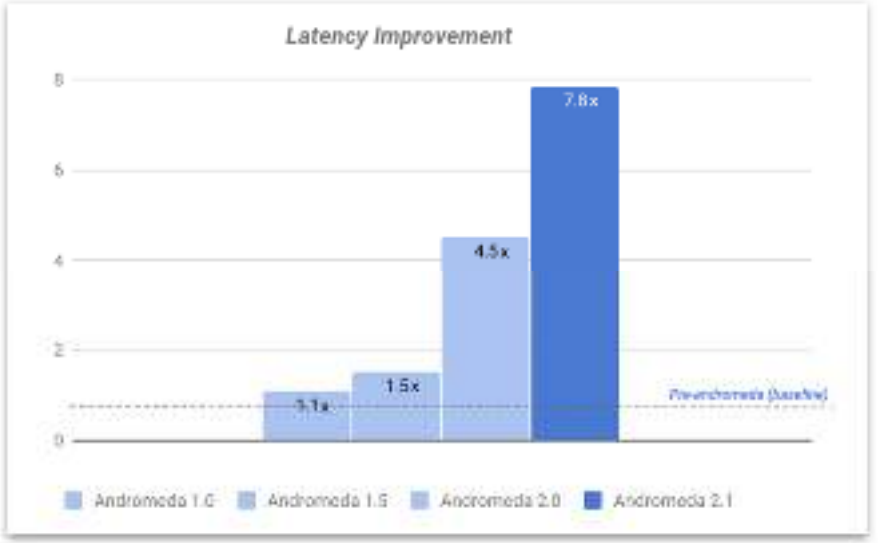
Google Networking

Performance

- 15,000 VMs per VPC
- Scalable bandwidth
  - 2 Gbps per vCPU
  - Up to **32 Gbps** per VM
  - Up to **100 Gbps** on GPU VMs
- Predictable, low latency (~10 µs)
- Best Practices for MPI workloads available (Google: “Google MPI Best Practices”)

Efficient SDN network

- Clos topology, a collection of smaller custom switches arranged to provide the properties of a much larger logical switch.
- Software Defined Networking allows us to improve latency with every release.



Google Cloud

Belnet

HPC VM Image

Google Cloud

HPC VM Image

- **CentOS 7 based VM Image**
- **Tunings Included**
  - Disable Hyper-Threading
  - Adjust user limits on system resources
  - Increase tcp *mem settings
  - Use the network-latency profile
  - Disable Linux firewalls
  - Disable SELinux
  - Intel MPI collective tunings
- **Will be offered through GCP marketplace**
- **User Guide, Alpha Sign Up Form**

Recently introduced features and tunings deliver major improvements for MPI/TCP

Compute-Optimized VMs

- C2 VMs use Intel Xeon 2nd Gen Scalable processors, can run at sustained 3.8 GHz.
- Fixed virtual-to-physical core mapping and *expose NUMA cell architecture to guest OS*

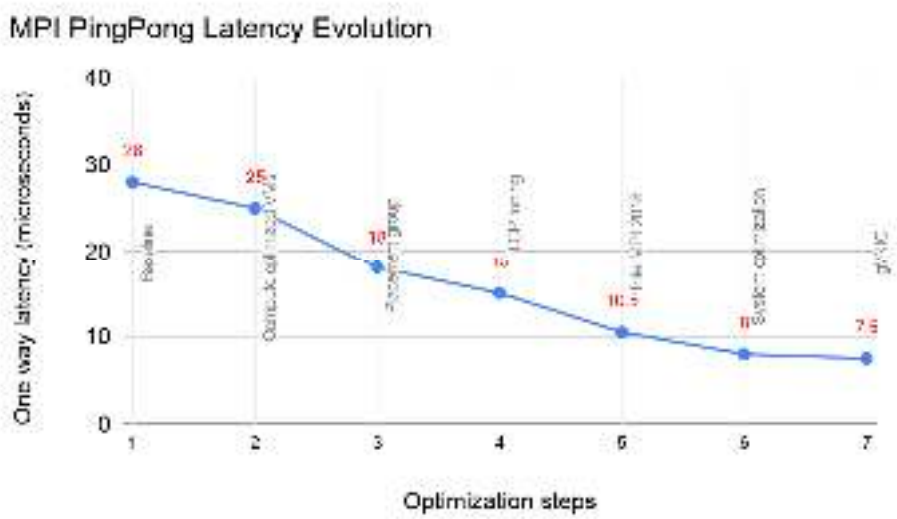
Compact Placement Policy

- Places VMs physically close in a single availability zone, reducing comm. latency
- Recently introduced support for *100+ VMs in a placement policy* (Limited Preview)

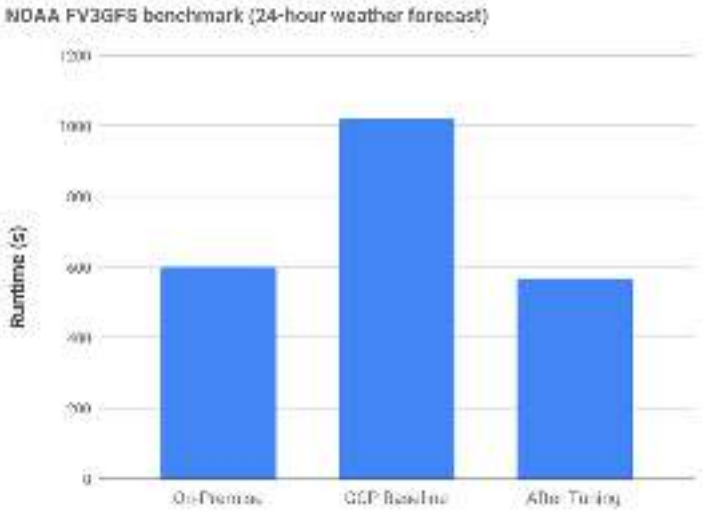
TCP/MPI Tunings

- Tunings and optimizations to *reduce latency for small messages to under 10us*
- *CentOS 7 HPC VM Image* helps create HPC-ready VMs out-of-box (Limited Preview)

Recently introduced features and tunings deliver major improvements for MPI/TCP



Between C2-60 instances with Compact Placement



With C768 model and 104 C2-60 Instances (3120 Cores)

Benefits of HPC VM Image

Quickly create HPC-ready VMs Out of box

- *Always stay up-to-date* with the latest GCP updates for tightly coupled workloads
- *No need to manually tune* MPI performance, no need for VM reboots

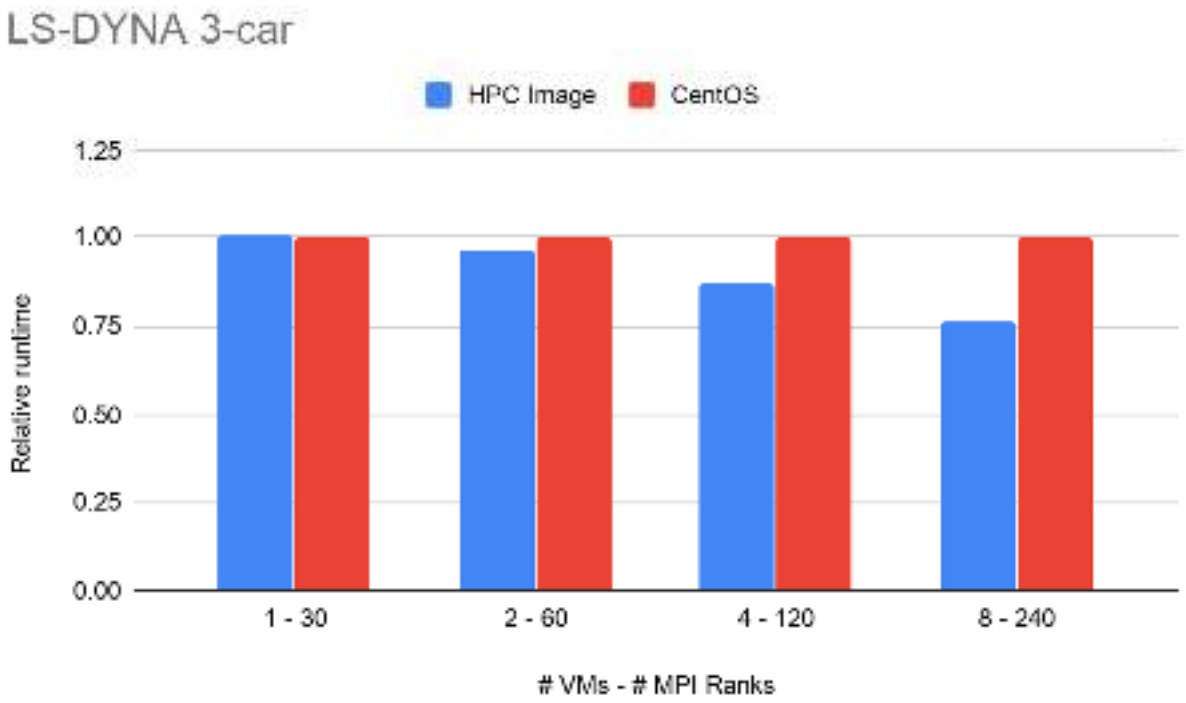
Tunings to unblock tightly coupled HPC workloads

- Tunings and optimizations to *reduce latency for small messages*
- Will benefit applications which heavily depend on collective communications

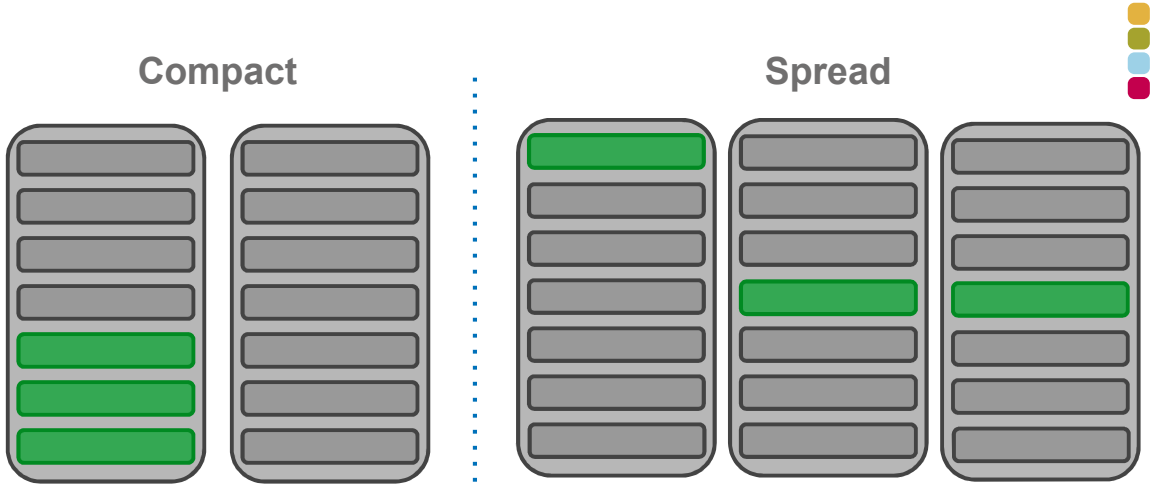
Consistent Performance, Improved Usability

- Achieve *consistent, reproducible application-level performance*
- Easier HPC *cluster setup and maintenance* for HPC system administrators

LS-DYNA simulation comparison using HPC VM Image



Reduce inter-node network latency with Placement Policy



instances placed within the same logical infrastructure, **lowering latency** between nodes.

instances will be placed on distinct hardware for **higher reliability** and reducing the impact of underlying hardware failures

- Spread Placement Policy available on General Purpose (N1, N2, N2D) and Compute-Optimized (C2) machine types
- **Compact Placement Policy** available on Compute-Optimized (C2) machine types
- The **maximum number** of instances per compact placement policy is 22
- Instances that use the same placement policy must be within the **same** availability zone

Google Cloud

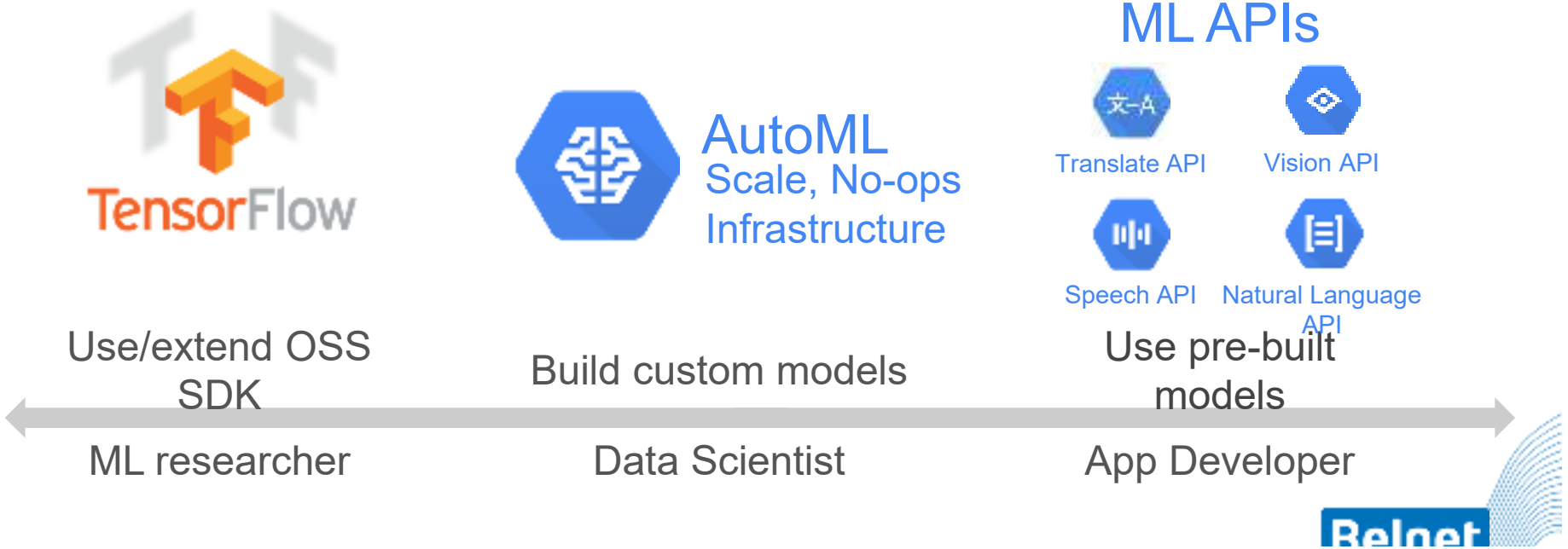
Belnet

Machine Learning

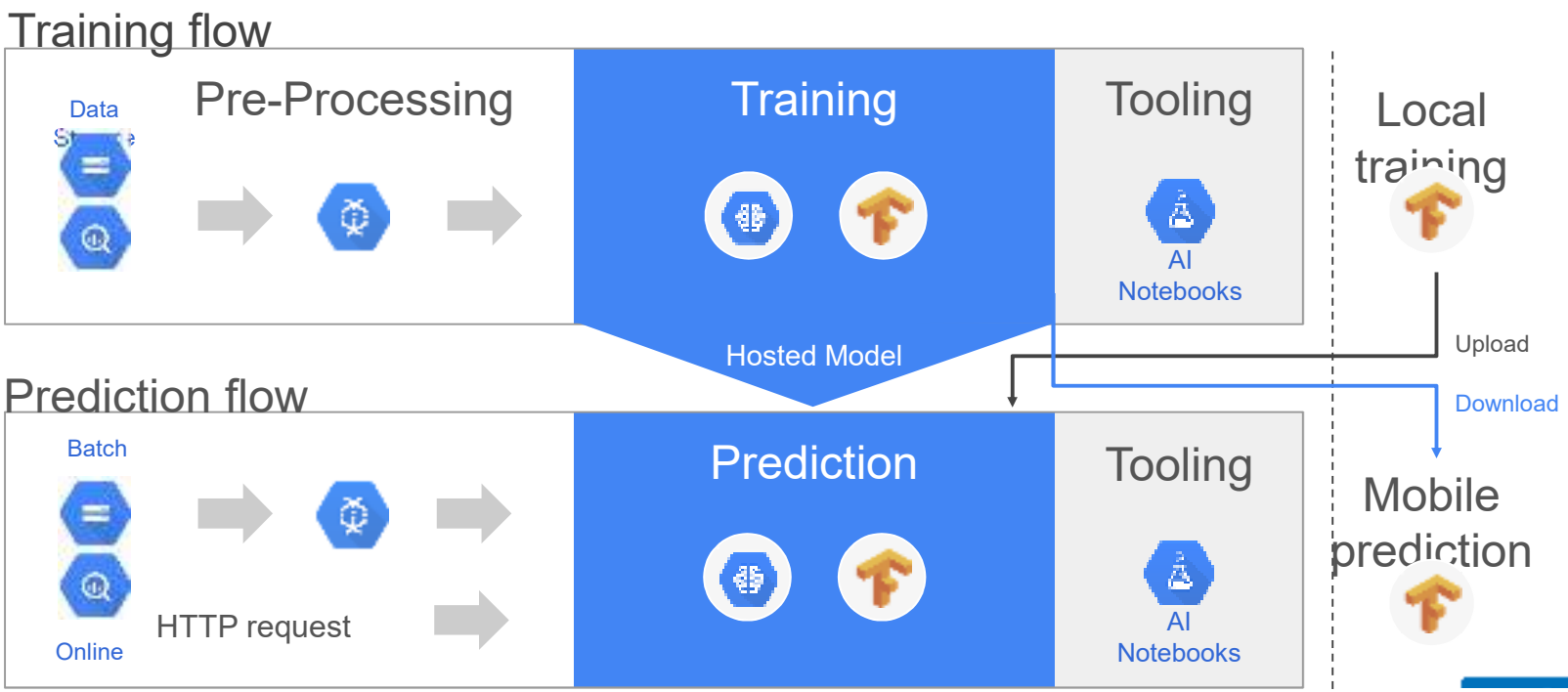
Google Cloud

Belnet
dedicated connectivity

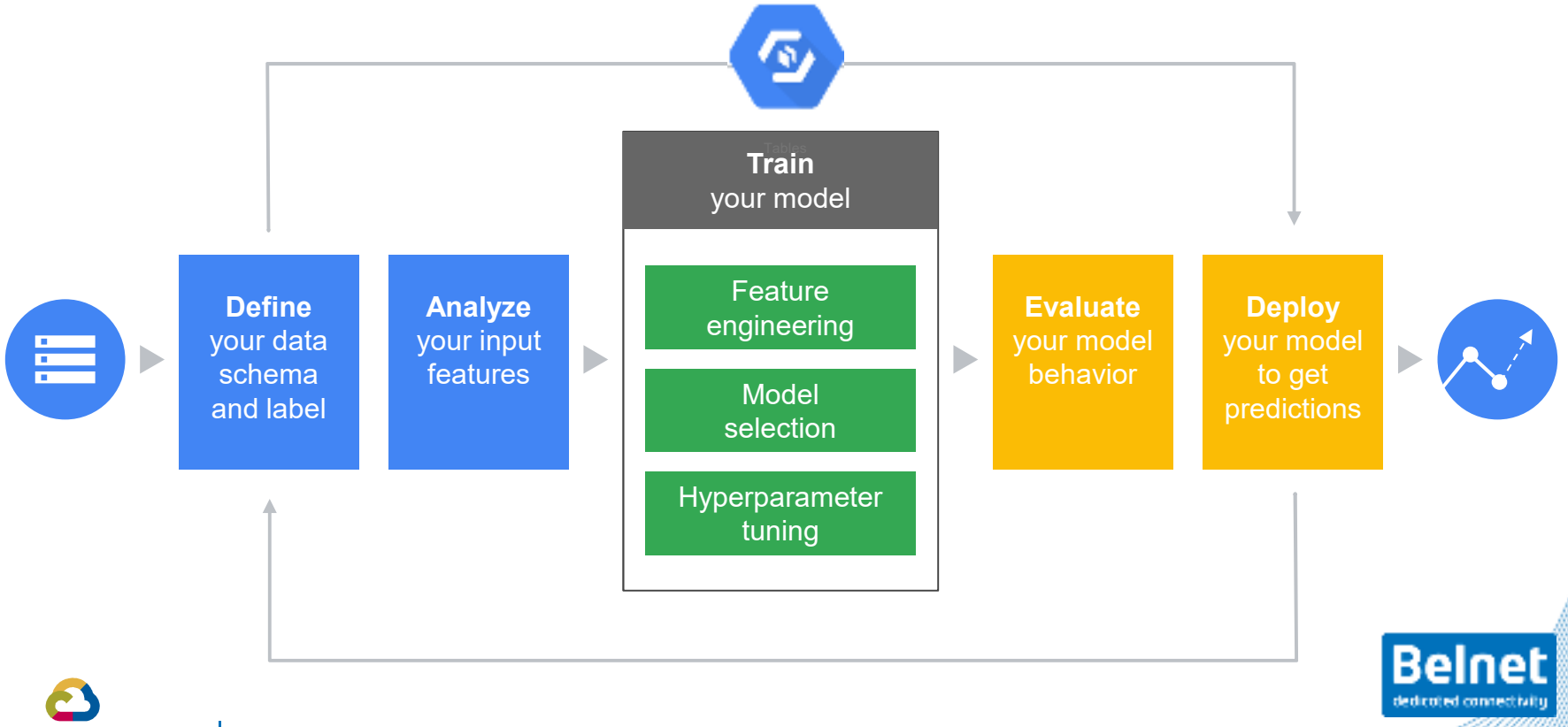
Offering across the spectrum



The flexibility of GCP



End-to-end Machine Learning lifecycle

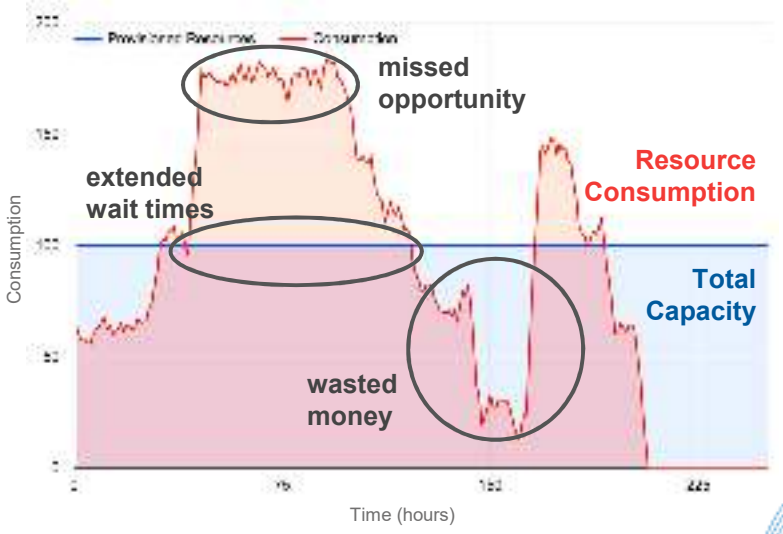


Optimizing utilization

bit.ly/low-cost-hpc

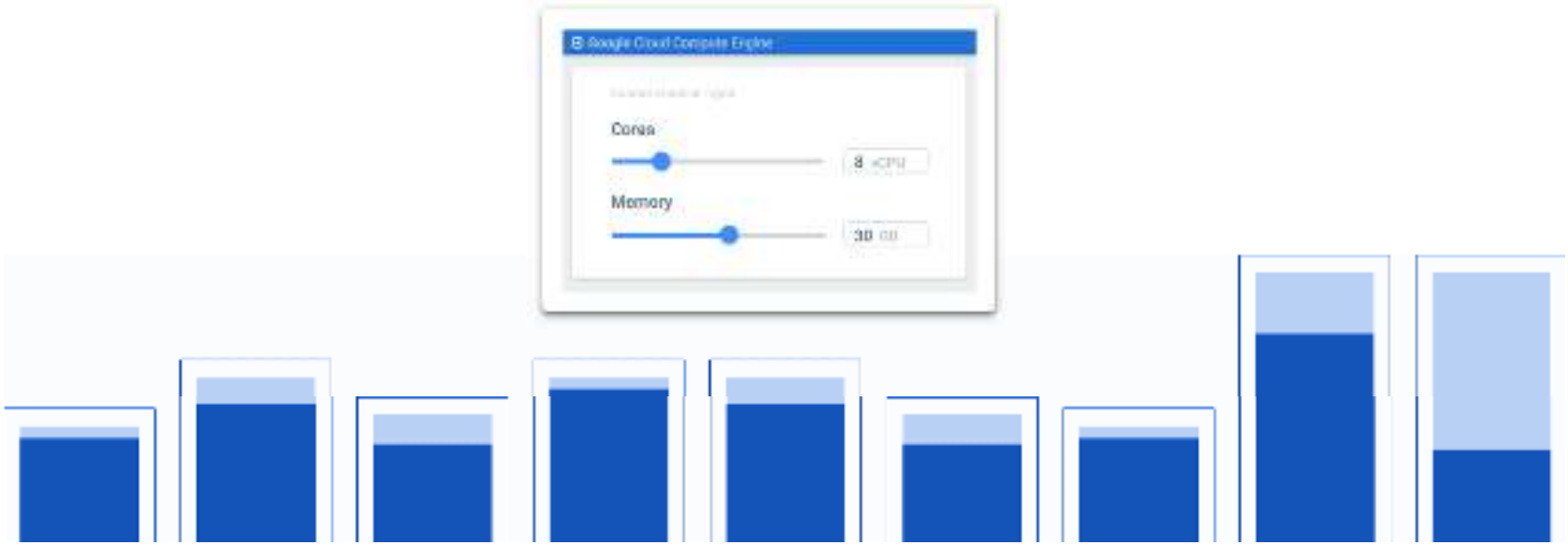
Optimizing cost and resource utilization

CPU, GPU, Storage, Network, Servers, Building, Utilities, Cooling



Custom Machine Types

Any CPU, Any Memory
Average **19%** Savings



Google Cloud

Belnet
dedicated connectivity

Preemptible VMs

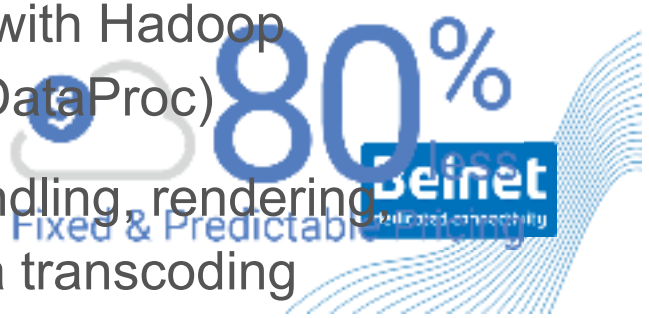
Made for batch, checkpointed, and high throughput computing.

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-

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



Cost Effective HPC

“We managed to get the cost of our main Best Practices analysis pipeline down from about ! As you can imagine that kind of cost reduction has a

per research dollar”

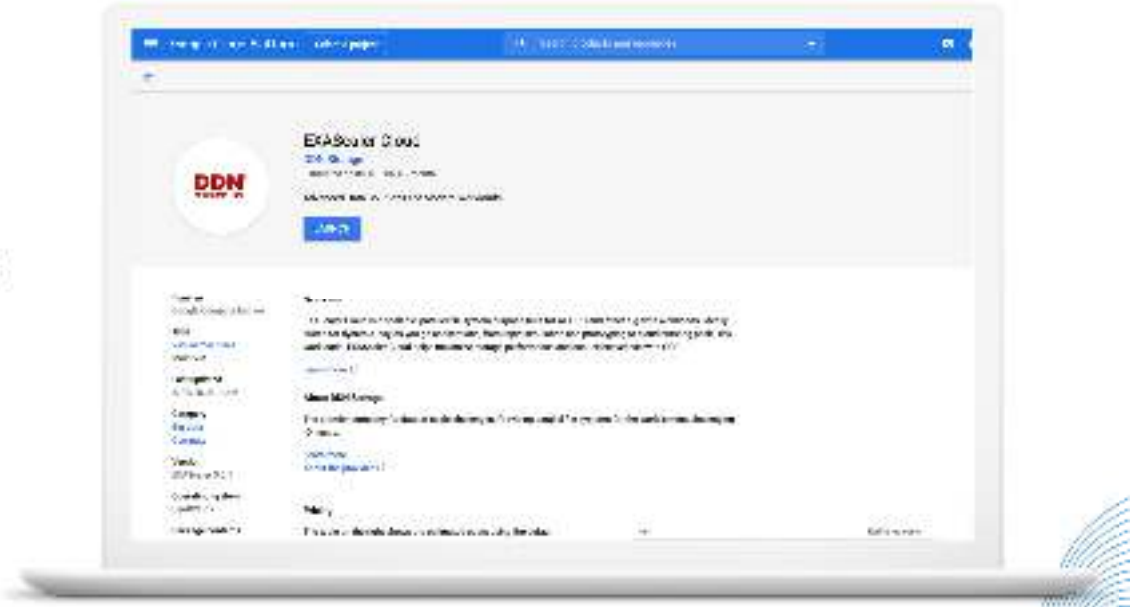
Google Cloud

Belnet
dedicated connectivity

HPC Software

DDN EXAScaler Lustre

- **DDN EXAScaler Cloud** is available through the **GCP Marketplace**
  - Includes **24x7 support**
  - Scalable, efficient, parallel file system purpose-built for HPC
  - Supports **Persistent Disk**



For more information, **visit the GCP Marketplace** or

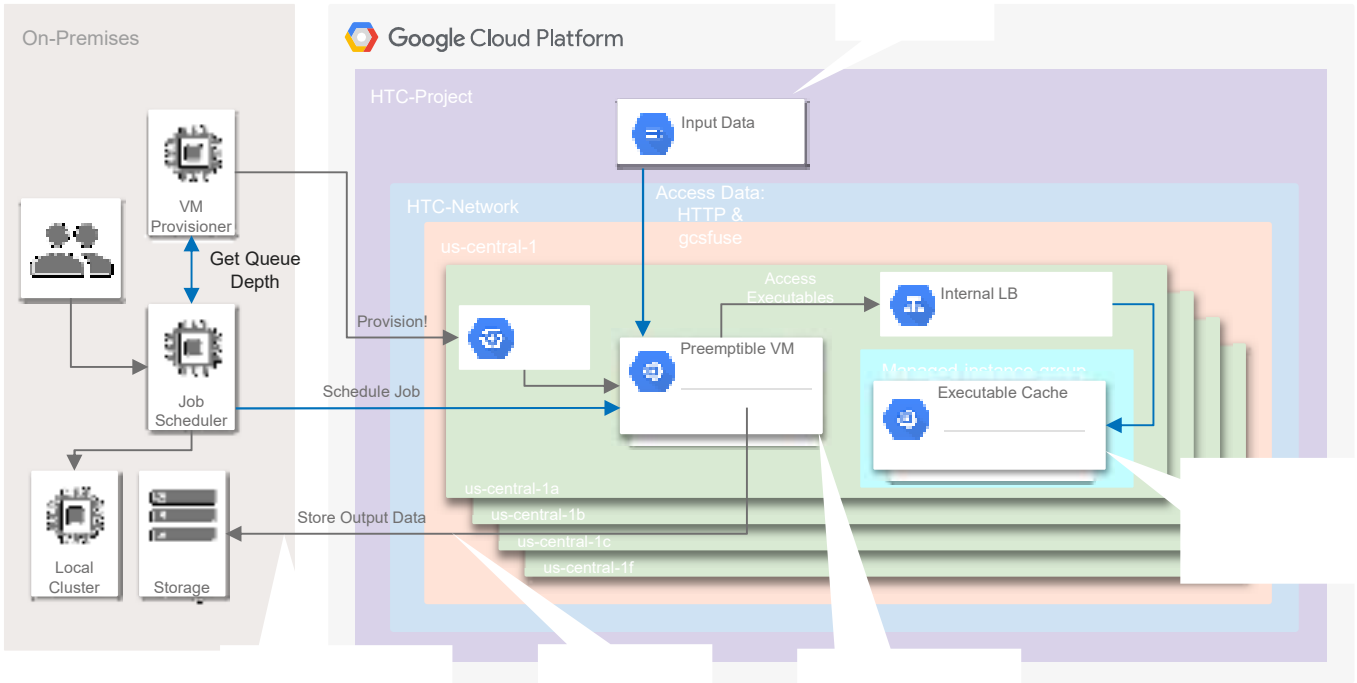
Slurm Workload Manager

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
- **Auto-Scaling Slurm Cluster Tutorial:** https://codelabs.developers.google.com/codelabs/hpc-slurm-on-gcp

Monte Carlo bursting on GCP via HEPCloud.



Genomics Pipelines API

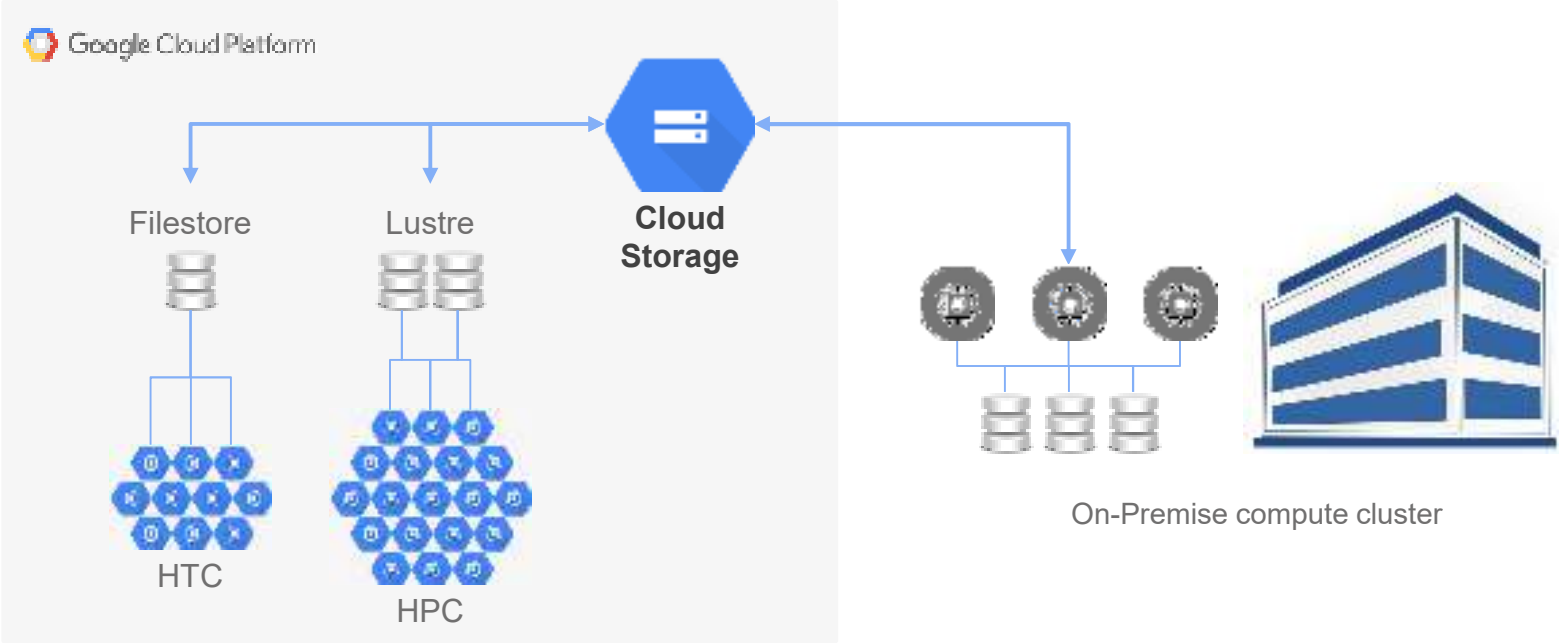
Batch scheduler for containers

Compose existing popular bioinformatics tools such as SAMtools, GATK, or DeepVariant into scalable pipelines. Utilize standard components packaged in Docker.

Multi-zone support, automatic quota management, queuing, input/output localization, and preemptible VMs support to remove the hassle of manual batch compute management.

Process thousands of samples with fully elastic multi-step pipelines without any cluster setup or management.

Hybrid storage model



Google Cloud

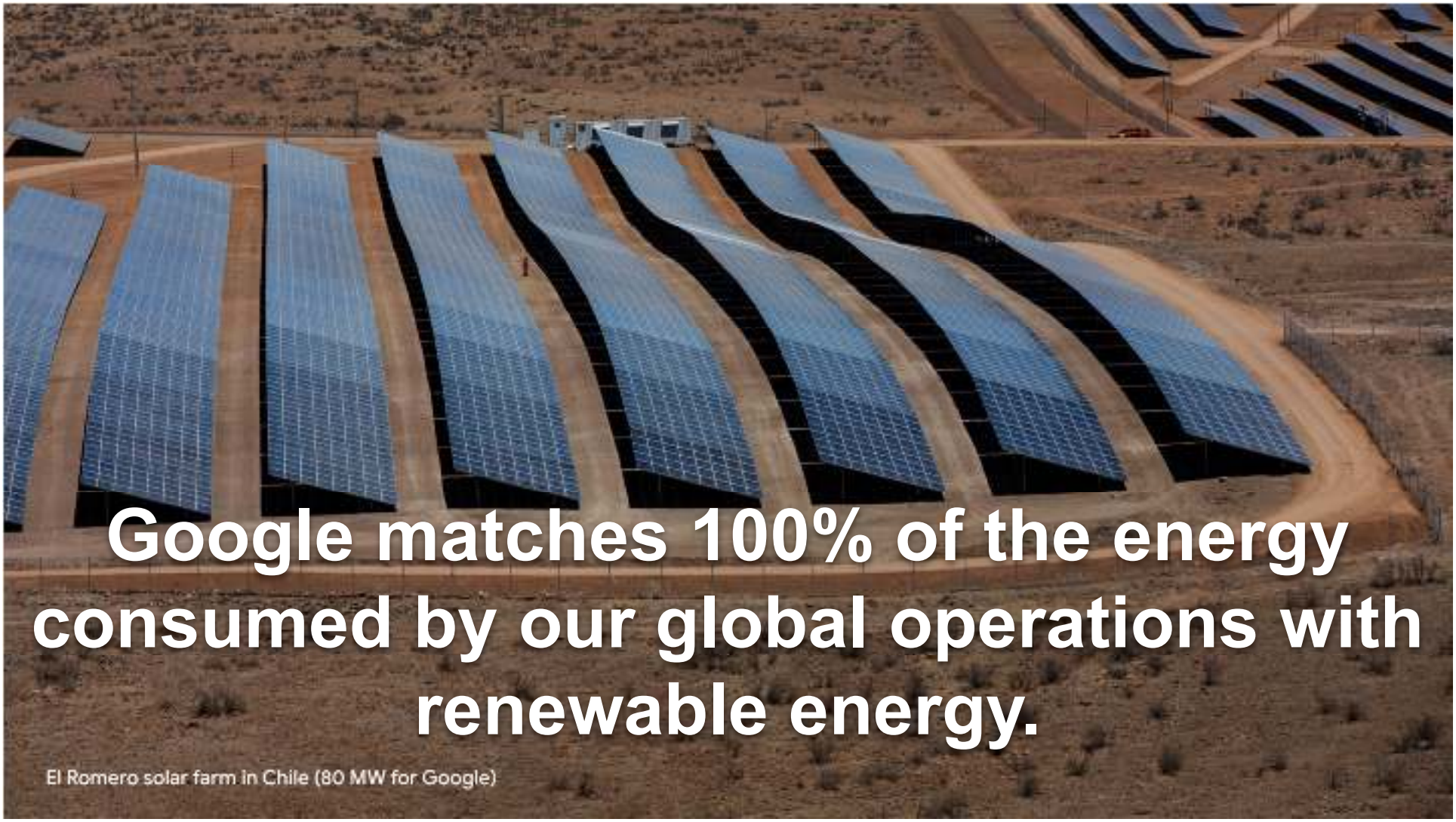
Belnet
dedicated connectivity

Google Public Datasets

https://cloud.google.com/public-datasets/

Belnet
dedicated connectivity

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


Google matches 100% of the energy consumed by our global operations with renewable energy.
El Romero solar farm in Chile (80 MW for Google)

GCP Credits for University Courses & Research

GCP Faculty Grants Program
Free credits for student learning & coursework

GCP Research Credits Program
Free credits for academic research workloads

Thank you.

https://cloud.google.com/hpc

Google Cloud

Belnet
dedicated connectivity

Feedback
&
Q&A

Belnet
dedicated connectivity


Thank you
for your attention
Belnet