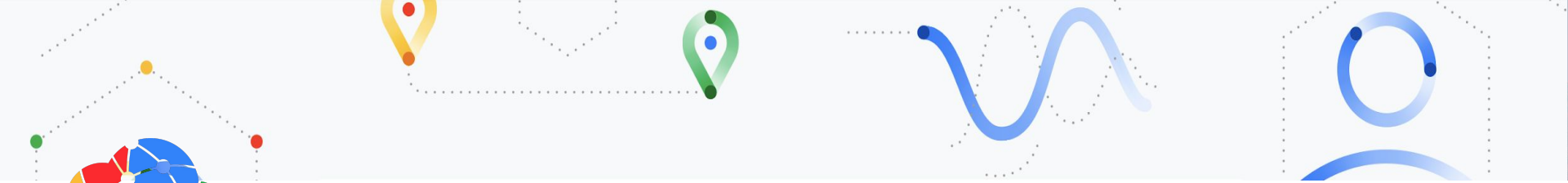


Google Cloud Networking Takeoff in RDU

Labcorp series

Sławomir Chyłek
March 29, 2022

Agenda

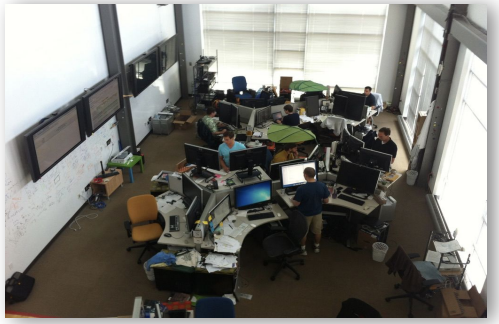
- Google Cloud in RDU
- Overview of Google Networking
- Opportunities Ahead
- Overview of Cloud Networking
- Q&A

Google Cloud in RDU

Google Cloud

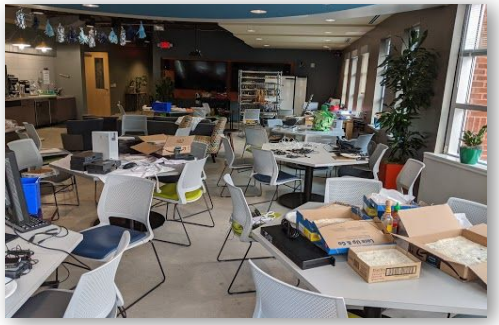
Google Growth in RTP

2005
4 Googlers *stealth mode*
(**Android**)



...2010
12 Googlers
Android

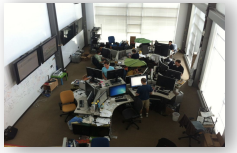
2011...
12 Googlers
Chrome

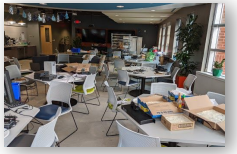


...2021
60+ Googlers
and counting

Google Cloud

1,000+ jobs over the next three-five years





60 jobs over in 15 years...

Hiring

Research Scientists

Administrative Business Partner

Technical Program Management

Solutions Consulting

Security/Privacy Engineering

Software Engineer

Staffing

Developer Relations

Site Reliability Engineer

Technical Solutions Consulting

Product Operations

Data Science

Product Manager

New Business Sales

Interns

RTP Growth with Google Cloud

[Growth]

Across North Carolina:

- 600 employees to date
- Chapel Hill office since 2005
- Data center in Lenoir County

Google Cloud in Durham

- 200 Morris Street
- **1,000+ jobs** over the next three-five years
- Top five Google Cloud engineering hubs in the U.S.

[Culture]

- Employee Resource Groups
- Well-Being Programs, like Yoga Meditation & Massage
- Community Engagement

Investments

Cloud Storage

Cloud Networking

Chief Information Security Office

Systems Infrastructure

Cloud Security

Cloud Apps

Site Reliability Engineering

Data Analytics

Overview of Google Networking

Google Cloud

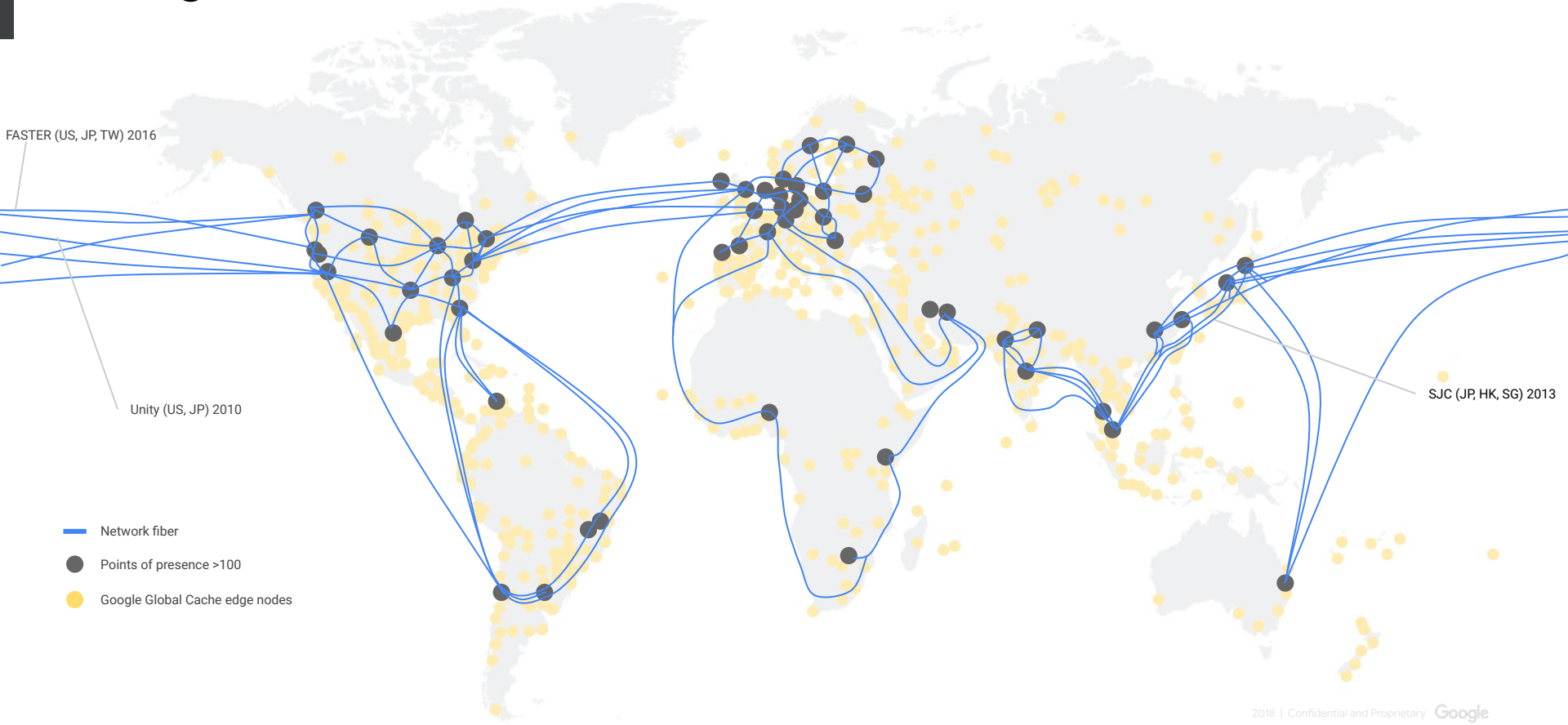
Three Drivers for Google Networking

Great performance and reliability to a global audience
Responsiveness is stickiness

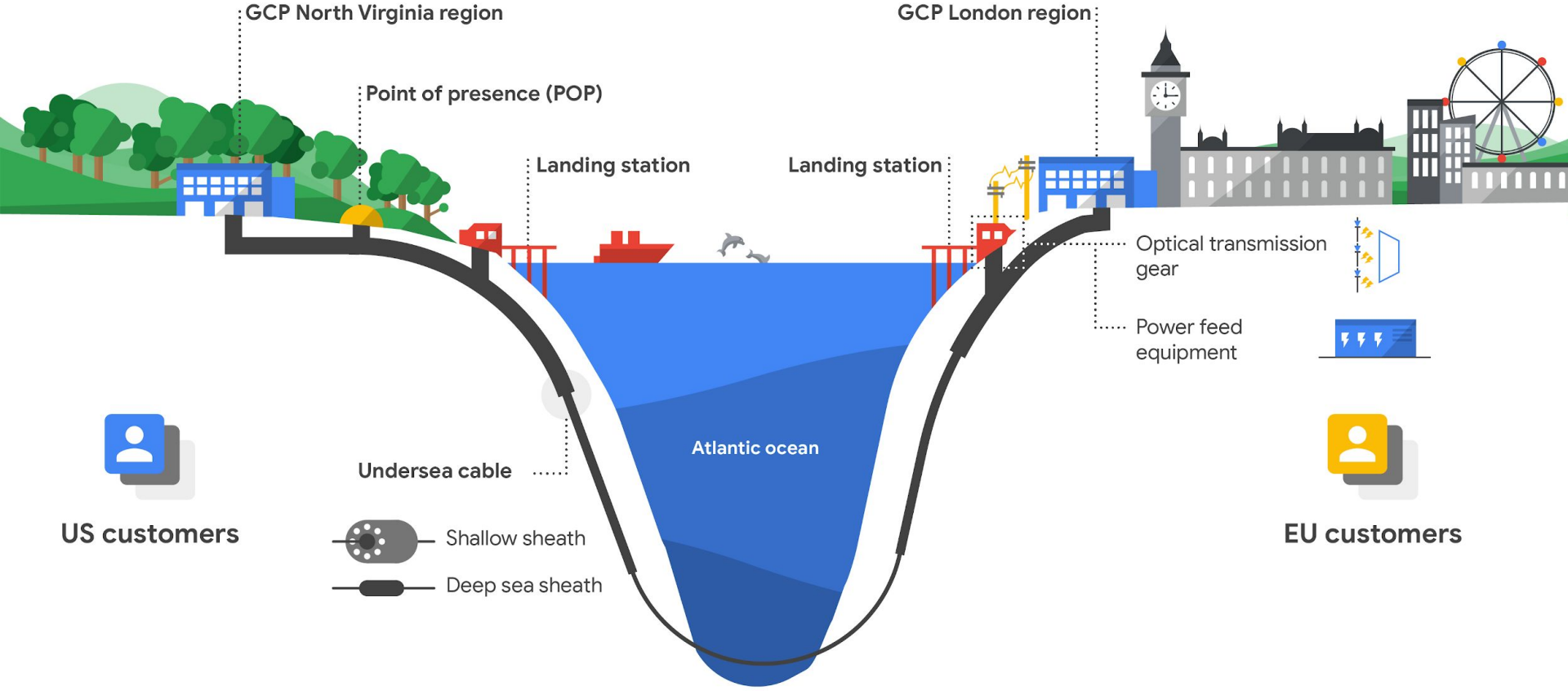
Seamless replication and failover across the planet
The whole is much greater than the sum of its parts

Revolutionary computing models
Storage and accelerators from thousands of servers as if local

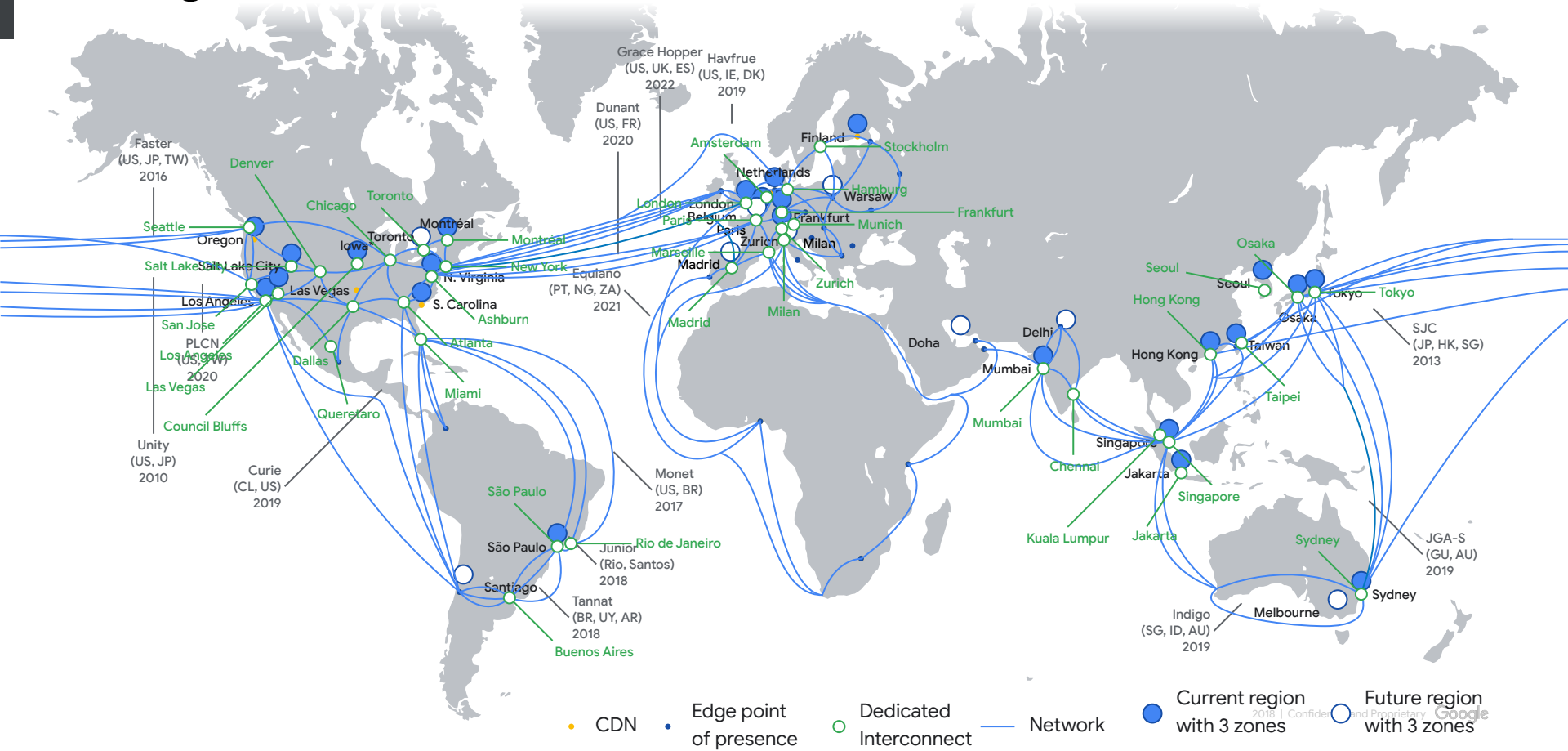
Google Network: Global Reach



Subsea network overview



Google's Global Network



Breadth of Google's networking publications

SDN networking in the real world:

- B4: Experience With a Globally-Deployed Software Defined WAN (SIGCOMM 2013)
- B4 and After: Managing Hierarchy, Partitioning, and Asymmetry for … (SIGCOMM 2018)
- Jupiter Rising: A Decade of Clos Topologies and Centralized Control in … (SIGCOMM 2015)
- Taking the Edge off with Espresso: Scale, Reliability and Programmability … (SIGCOMM 2017)

Congestion, traffic management, and load balancing:

- BBR: Congestion-Based Congestion Control (CACM 2017)
- Bandwidth Enforcer: Flexible, Hierarchical Bandwidth Allocation for WAN … (SIGCOMM 2015)
- TIMELY: RTT-based Congestion Control for the Datacenter (SIGCOMM 2015)
- Maglev: A Fast and Reliable Software Network Load Balancer (NSDI 2016)
- An Internet-Wide Analysis of Traffic Policing (SIGCOMM 2016)
- The QUIC Transport Protocol: Design and Internet-Scale Deployment (SIGCOMM 2017)
- Credit-scheduled Delay-bounded Congestion Control for Datacenters (SIGCOMM 2017)

See http://g.co/research/networks

...and more breadth

End-host networking:

- Andromeda: Performance, Isolation, and Velocity at Scale in Cloud ... (NSDI 2018)
- Carousel: Scalable Traffic Shaping at End-Hosts (SIGCOMM 2017)
- Eiffel: Efficient and Flexible Software Packet Scheduling (NSDI 2019)

Design and management at scale:

- Evolve or Die: High-Availability Design Principles Drawn from Failures in a ... (SIGCOMM 2016)
- Libra: Divide and Conquer to Verify Forwarding Tables in Huge Networks (NSDI 2014)
- Condor: Better Topologies through Declarative Design (SIGCOMM 2015)
- Minimal Rewiring: Efficient Live Expansion for Clos Data Center Networks (NSDI 2019)
- SIMON: A Simple and Scalable Method for Sensing, Inference and Measurement ... (NSDI 2019)

See http://g.co/research/networks

Opportunities Ahead

Google Cloud

1) Converging Compute, Storage, and Networking

Google Cloud

What We Should Look Forward To

- Rethinking the basic architecture for interactive serving systems
- Rethinking granularity of load balancing decisions
- Rethinking function-shipping versus data shipping

Haven’t We Solved All of These Problems Already?

- Well, yes and no.
- But, we are good at performance work and we are good at building the simplest possible system to demonstrate the good ideas
- What would it mean to define the “simplest complex” distributed systems scenario together? What would it mean to optimize it?
- How can we break outs of the “networking” box into end to end systems?

Paths Forward for the Next Order of Magnitude in Compute Performance

- Huge fraction of server CPU consumed by communication overhead
- OS Scheduling not built for low-latency communication
- Network fabrics are embarrassingly parallel

2) Tight Coupling → Great Expectations

Networking 101

- *Best Effort:* no guarantees about whether data will arrive
  → What are the implications for performance, predictability, and availability?

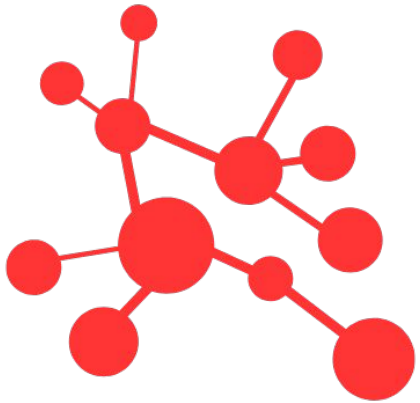
- *Soft State:* useful for efficiency but not required for correct operation
  → What are the implications for performance and predictability?

What is the Network Product Definition?

- **Best effort** and **soft state** as the law of the land?
  - But how do you build services around that?
- Service-specific availability of a given capacity... but not more?
- Latency predictability and stability is absolutely critical
- All based on consistent, meaningful measures
  - Tied to higher-level measures of application requirements

Software Defined Networking

- Deep understanding of scalable, fault tolerant, logically centralized services
  → Spectrum between soft state and hard state
- Centralized control and admission control → Predictability
- Requirements for virtualization and global policy
- Best effort → SLOs and SLIs



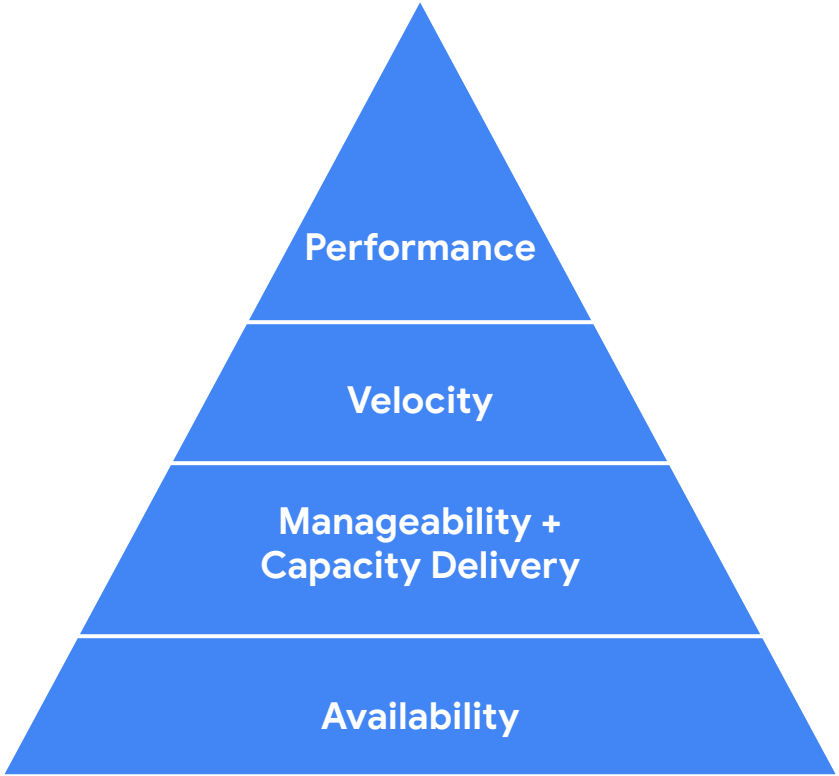
Raising the Level of Abstraction

- Developers do not care about IP packets or TCP connections

- Entry point is the Remote Procedure Call

3) Infrastructure Work

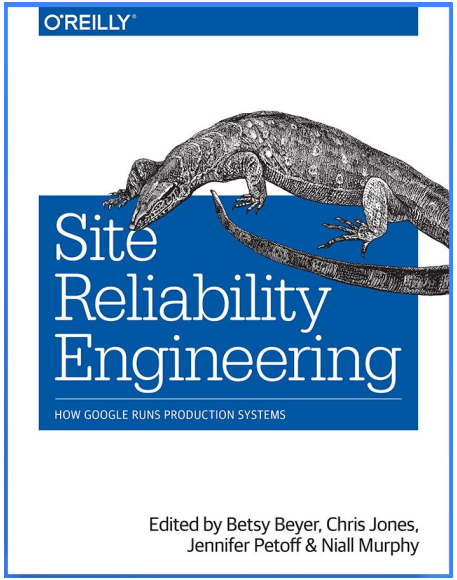
How we Prioritize Infrastructure Work



Availability is Paramount

- First things first: an insecure infrastructure is an unavailable infrastructure
- Stability is more important than efficiency
- Network management is critical
- Configuration is hard
- Automation matters but can be counter to availability

"Evolve or Die: High-Availability Design Principles Drawn from Google's Network Infrastructure." SIGCOMM 2016.



Build for Velocity

- Velocity is the speed of iteration
- Build for hitless upgrades and self-validation
- Debugging and tracing matter
    - Without visibility, performance does not matter
- Launch and iterate

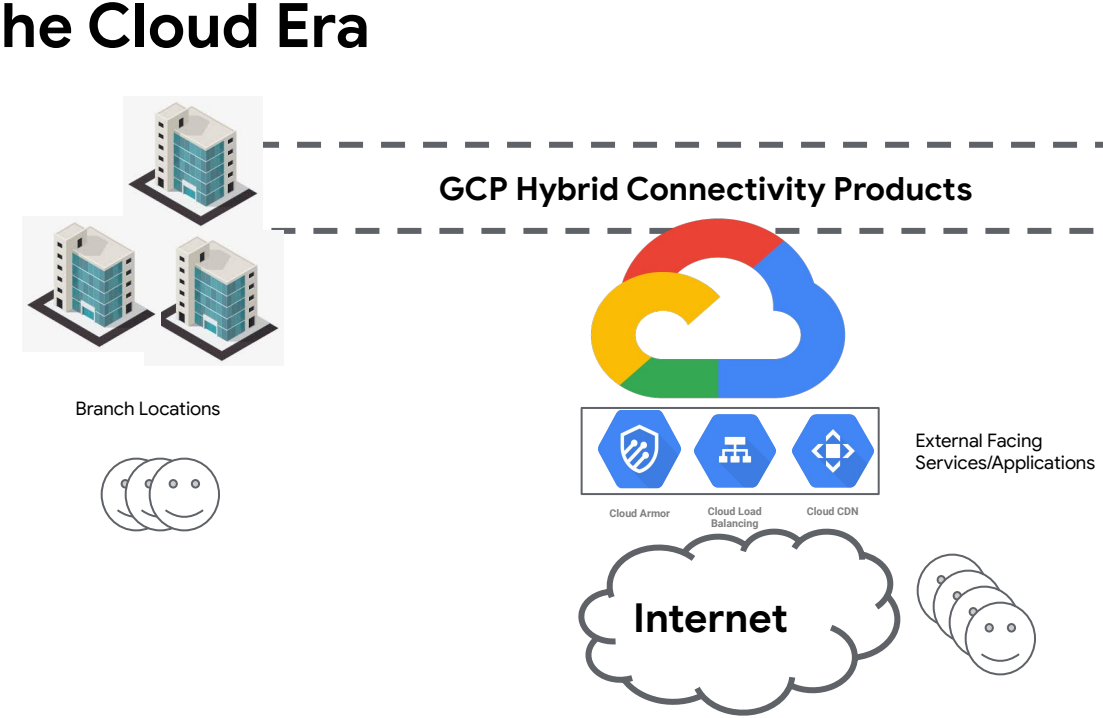
Google Cloud Networking

Google Cloud

Cloud Networking Mission

Seamlessly and securely connect and serve any application, any user, anywhere

Opportunity to be the center of the Enterprise Network in the Cloud Era



Why Google?

- Global Reach
- High-Performance Backbone
- Open Strategy (i.e. Anthos)

Customer-Value: Uplevel the customer's view from Routes/Interconnects to Services + Policy. Move away from special-purpose hardware for security and weave this into the fabric of Cloud. Allow them to focus on what matters.

Bringing 4-Key Strategic Areas to NC

Hybrid Connectivity

*"Meet our customers where they are, enable them to build and deploy applications globally in multi-cloud and off-cloud environments, and be the **connectivity provider of choice**."*

- Network Connectivity Center
- Multi-Cloud Connectivity
- Cloud Egress Products

Network Security

*Make **Network Security invisible** by weaving it into the cloud fabric with automated verifiable protection, simplified control, and only alert on sophisticated threats with relevant context*

- Secure Cloud Fabric
- Advanced Firewall
- Secure Web Gateway

Load Balancing

Enable users to build the most scalable, dependable and highest-performing applications and services on all platforms and across all clouds.

- Network Load Balancer

Service Directory

Enable enterprises to publish, discover, and connect multi-cloud services regardless of their environment.

- Service Directory

Thank you!