Things to read while you wait:

• **Creating an effective poster**, John Wilkes
  [http://goo.gl/HUVffd](http://goo.gl/HUVffd)

• **The 4 sentence abstract**, Kent Beck
  [https://goo.gl/W5SCMT](https://goo.gl/W5SCMT)
Building warehouse-scale computers or ... what's it like to supply exponential growth

john wilkes  2018-11
Scale has been the single most important force driving changes in system software over the last decade.

– Technical perspective: Is scale your enemy, or is scale your friend?
You’re all thinking too small.
Some of you are thinking too small.
2018-11-05 1PB scan

```
SELECT *
FROM `google.com:bigquery-petabyte.retail_petabyte.value_partitioned`
WHERE customerKey = "1104796800000-19548818"
```

Query results

Query complete (5 min 49.309 sec elapsed, 1.09 PB processed)

 elapsed time
5 min 49.309 sec

Slot time consumed
5 days 16 hr

Worker timing
Here's the system!
What if …
the “**system**” included not just the computers, but the network fabric and the WAN endpoints.
and the cooling system
and the building-management system
and the power system
and the technicians
In fact ... the entire warehouse scale computer
By the way … it costs O($200M). No pressure.
Key challenges

Add capacity optimally

The Fleet

Keep capacity optimized
Planning for compute resources

- CPU
- RAM
- disk, SSD – capacity, performance

- new NVRAM thingies ...
- accelerators (complication: a wide variety)
- inter- and intra-datacenter networking
- power
- datacenters
- land, water, sewage
- ...

All have multiple generations/options
Response times are weeks/months/years, not milliseconds

http://googleasiapacific.blogspot.se/2015/06/growing-our-data-center-in-singapore.ht
Google Network

The largest cloud network, comprised of more than 100 points of presence
How long does it take to build capacity?

Note: variance is as bad as delay
Planning for compute resources

What kind of machine to buy?
- different groups want different things (e.g., search, Cloud)

⇒ MotD + customizations

Idea: reuse machines when they get handed down
- when is it worthwhile? (price of power? room for expansion?)

Also: do we have ... space? power? networking? budget?
- models; what-if analyses; uncertainty (demand, supply)
- objective function: total cost-of-ownership
Planning for compute resources: Total Cost of Ownership (TCO)

what do you mean by “total”?  
• average? over what? continent? time?  
• depreciation schedule?

what do you mean by “cost”?  
• initial purchase price, or the average over time?  
• is delivery or installation included? what if they are being reused?

what do you mean by “ownership”?  
• machines can be transferred/given/sold/break  
• who “owns” a machine running a shared service for a customer?
A simplified overview

- **forecast**
- **time**
- **sites, data centers, power**
- **network, machines, storage**
- **usage**
- **$$ prices**
- **compute/storage capacity**
In a world of exponential demand growth

Traffic generated by servers in our datacenters
In a world of exponential demand growth
In a world of exponential demand growth

“Prediction is very difficult, especially about the future”
— Mark Twain(?) via Neils Bohr
A few factors affecting forecasts

- Better software
- Better hardware (e.g., accelerators)
- New features
- User growth

(time)
Putting it all together

- 1 second
- 1 minute
- 1 hour
- 1 day
- 1 week
- 1 month
- 1 year
- 2 years
- 5 years

Forecasts

resource demand

- Borg resource reclamation
- online workload prediction
- short-term PA demand forecasts mid-tiers
- long-term fleet forecasts
How much capacity do we need? and when do you need to know?
It takes a few moving parts ...
Cloud Datastore transactions/s

1X
Target Traffic

5X
Worst Case Estimate

50X
Actual Traffic

Original Launch Target
Estimated Worst Case
Actual Traffic
Planning for power

Continuous PUE Improvement
Average PUE for all data centers

PUE (Power Usage Effectiveness) = total Watts / compute Watts
- smaller is better
- industry average ~1.7
Planning for power

https://www.google.com/about/datacenters/efficiency/internal/
Planning for power

Some of the challenges

• Narrow range of experience
• Exploration may discover unsafe states
• Inputs out of our control (e.g., weather)
• Control system reliability/availability
• Agility (new hardware)
• Reinforcement learning with long delays between action and change in system state

• Safety. Safety. Safety.
We reached 100% renewable energy balance in 2017
Every hour of electricity use at Iowa data center

Although our Iowa data center achieved 100% carbon-free energy during the majority of hours in 2017, there is also a recurring reliance on carbon-based power — most notably in late summer, when wind speeds decline.

Overall in 2017, 74% of this data center’s electricity use was matched on an hourly basis with carbon-free sources.
Meanwhile – what’s up with Moore’s law?

Graph from 40 Years of Microprocessor Trend Data, Karl Rupp, CC-BY 4.0.
Meanwhile – what’s up with Moore’s law?

2015: Tensor Processing Unit (TPU)

2017: Cloud TPU
180 teraflops of computation
64 GB of HBM memory
Meanwhile – what’s up with Moore’s law?

2017: TPUv2: Tensor Processing Unit pods
11.5 petaflops; 4 TB HBM
2-D toroidal mesh network
Meanwhile – what’s up with Moore’s law?
Meanwhile – what’s up with Moore’s law?

2018: TPUv3
> 100 petaflops!
> 8x the performance of a TPUv2 pod
Putting it all together

Operations

Usage

Demand forecasting

Monitor + report

Delivery + fulfillment

Supply

Supply requests (orders)

Planning

Demand
Putting it all together – a few more details

Operations

Fulfillment

Supply (current + future)

Usage data

Historic-demand system

Demand forecast system

Customer demand system

Monitor + report

Supply planning

Ordering

Monitoring

Datacenters, power, parts

High-level services + lower-level stuff

Borg, Colossus + lower-level stuff

Machines, racks, ...

High-level services + lower-level stuff

Borg, Colossus + lower-level stuff

Machines, racks, ...

Datacenters, power, parts
Putting it all together – a few more details

- Operations
- Fulfillment
- Supply (current + future)
- Monitor + report
- Supply planning
- Policies
- Ordering

Datacenters, power, parts

Historic-demand system

Demand forecast system

Customer demand system

Usage data

Monitoring

High-level services + lower-level stuff

Borg, Colossus + lower-level stuff

Machines, racks, ...

Datacenters, power, parts
Putting it all together

Operations

 Demand forecasting

Usage

Monitor + report

Delivery + fulfillment

Supply requests (orders)

Supply

Planning

Demand
Putting it all together

0-6 months machines

Product Areas

Operations

Supply

Monitor + report

Planning

Policies

one Google-wide supply

each PA does these
Putting it all together – multiple timelines

- **Product Areas**
  - **0-6 months** machines
  - **12-18 months** parts
  - **24+ months** power, data centers

- **Operations**
  - Demand forecasting
  - Supply
  - Policies

- Supply forecast
2018 CapEx = $25.1B

(+$2.4B for an office building in New York)

source: Alphabet SEC filing

3-year trailing CapEx, as of March 2017
Final thoughts

• There’s a lot of technology behind “the cloud”
• At scale, efficiency *really* matters
• [and: we’re hiring!]