

Day7 - Cloud

02476 Machine Learning Operations

Nicki Skafte Detlefsen, Associate Professor, DTU Compute

January 2026

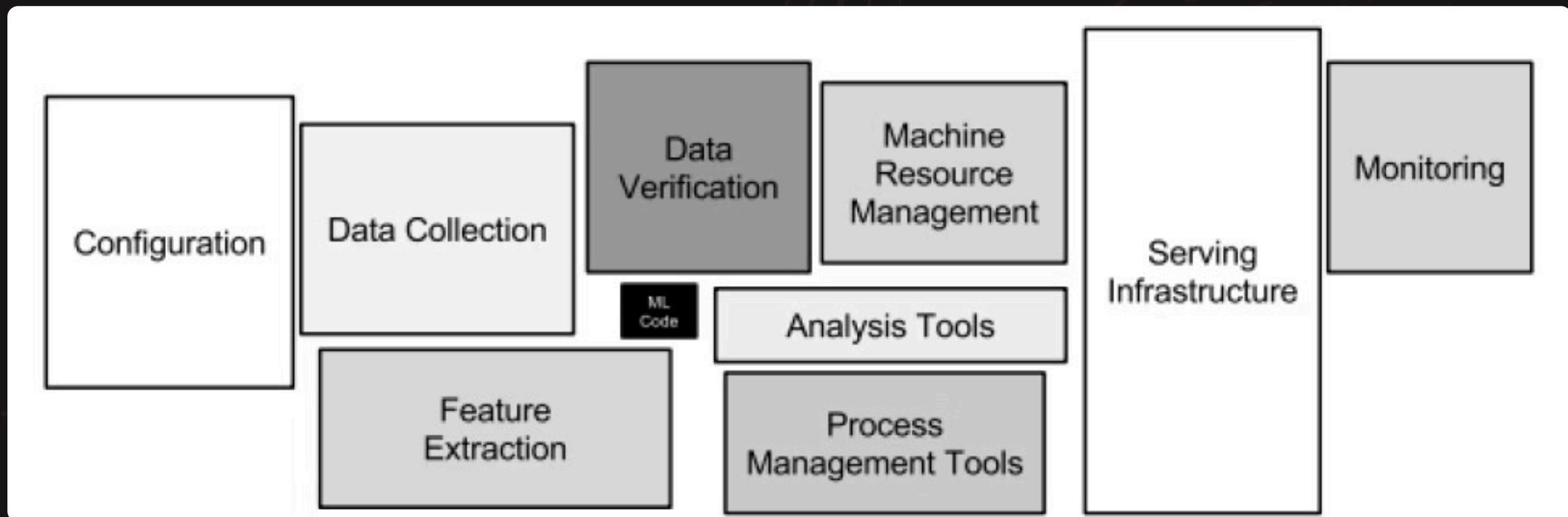
Conceptually: What is the cloud

It is to data scientists what advance calculus was to mathematicians in the early 17th century: The power of infinity



Technical debt in ML

A fraction of real world ML systems is composed of ML code, the rest is infrastructure.



The cloud provides services to help lower the technical debt

Why use cloud?

Reliability

Hardware failures, network outages, and system crashes require 24/7 monitoring and rapid response capabilities

Scalability

Predicting capacity needs and provisioning hardware months in advance creates inflexibility and wasted resources

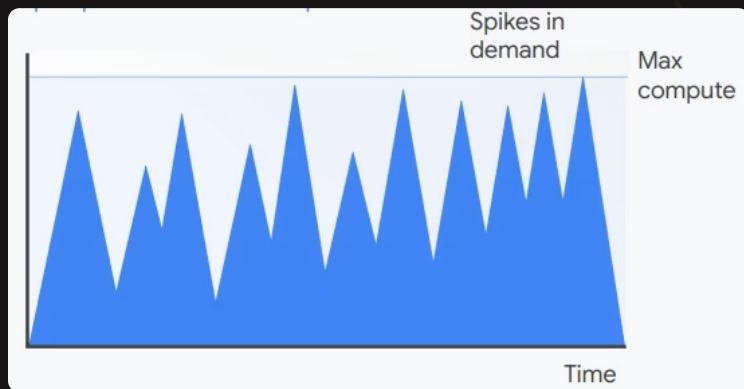
Complexity

Modern infrastructure stacks involve networking, storage, compute, security, and orchestration—each requiring deep expertise

Expertise

Hiring and retaining infrastructure specialists is expensive and challenging, especially for smaller organizations

Why cloud exists?



1. All teams will henceforth expose their data and functionality through service interfaces
2. Teams must communicate with each other through these interfaces
3. There will be no other form of interprocess communication allowed: no direct linking, no direct reads of another team's data store, no shared-memory model, no back-doors whatsoever
4. It doesn't matter what technology they use—HTTP, Corba, Pubsub, custom protocols
5. All service interfaces, without exception, must be designed from the ground up to be externalizable to developers in the outside world
6. Anyone who doesn't do this will be fired

2002, Jeff Bezos

It is this mindset of providing a lot of services that basically makes up cloud

Pros and Cons of using cloud

Benefits

- **Cost Savings:** Pay-as-you-go, no upfront hardware costs.
- **Global Accessibility:** Deploy worldwide, bringing services closer to users.
- **Scalability:** Instantly scale resources up or down with demand.
- **Unlimited Storage:** Store vast amounts of data without physical limits.
- **Automated Services:** Managed services handle backups, updates, and maintenance.

Challenges

- **Data Control:** Third-party infrastructure raises privacy and compliance concerns.
- **Internet Dependency:** Requires stable internet; outages can be critical.
- **Vendor Lock-in:** Difficult to migrate due to proprietary services.
- **Security Concerns:** Shared infrastructure increases attack surface.
- **Cost Management:** Without governance, costs can spiral out of control.

What is the cloud?

**Me: telling grandma what
is this GCP that I work on**

**Grandma: oh, so you're
renting computers?**

**You just insulted
my entire industry**

But yes.

In practice

1. Big data centers of interconnected computers.
2. A software stack on top



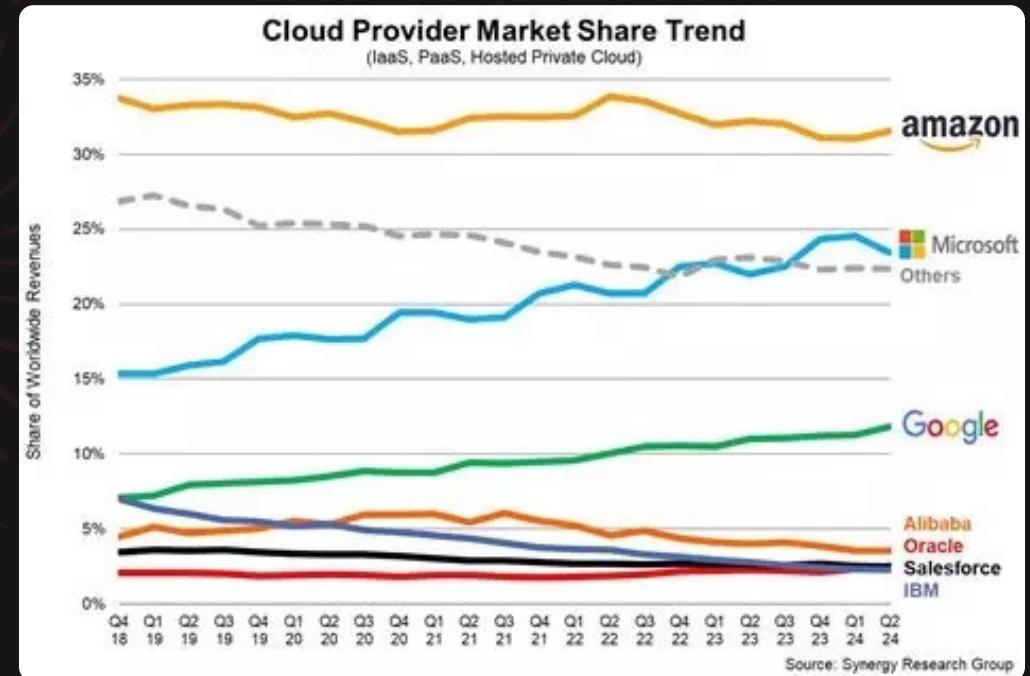
Cloud vendors

AWS are still the largest player in the cloud market.

However the others are growing faster.

We are going to take a look at GCP.

(because they gave me credits)



What does the cloud consist of

Each cloud vendor have a number of services

Depending on your application, only a subset is of interest



<https://cloud.google.com/products>

Exercise: what do you need?

Based on your current knowledge about machine learning and what you have already learned in this course, what kind of services would you need from a cloud provider to do Machine Learning at Scale?

Important services for ML



Engine



Bucket



Build



Registry



Vertex

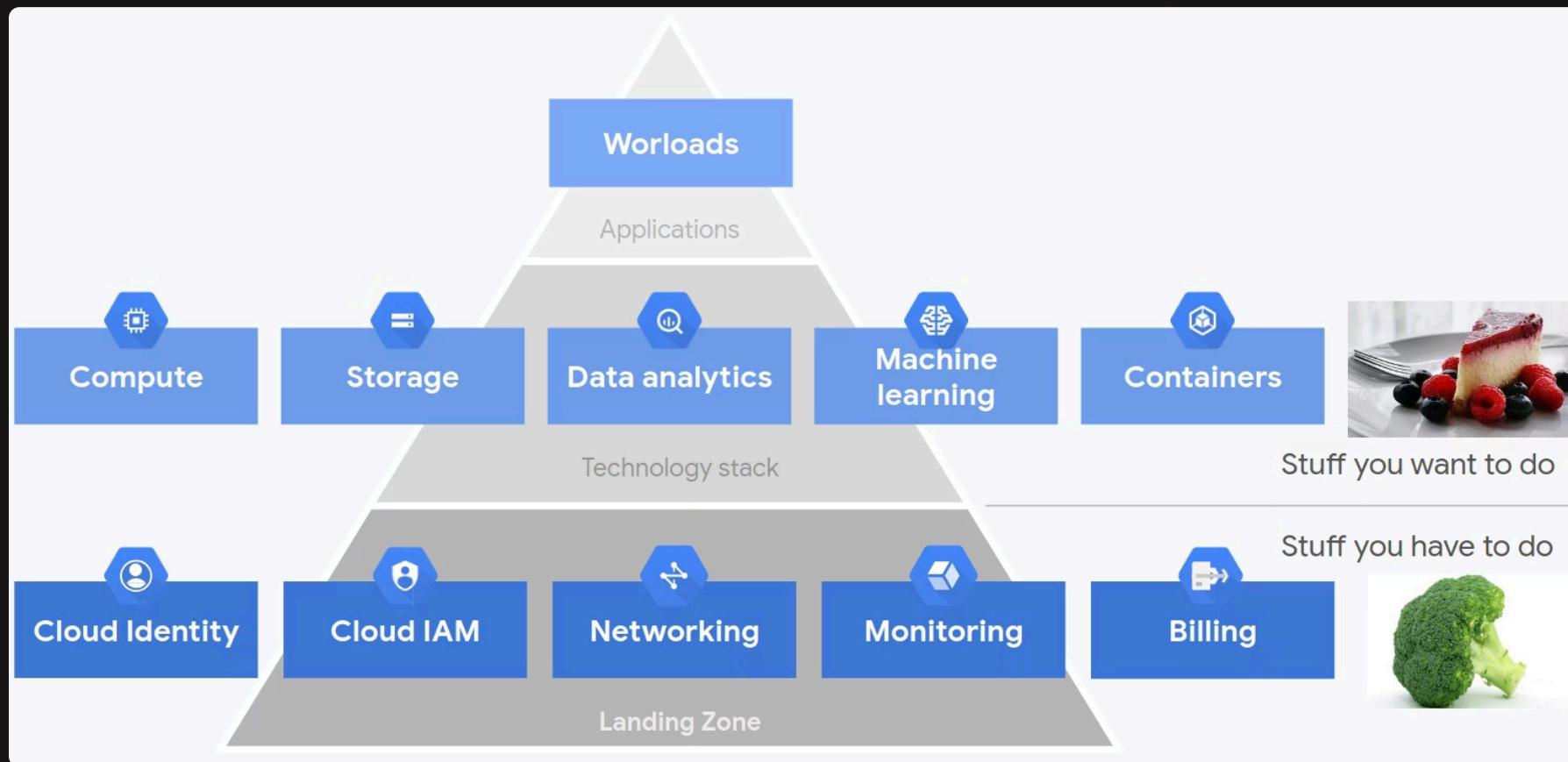


Functions

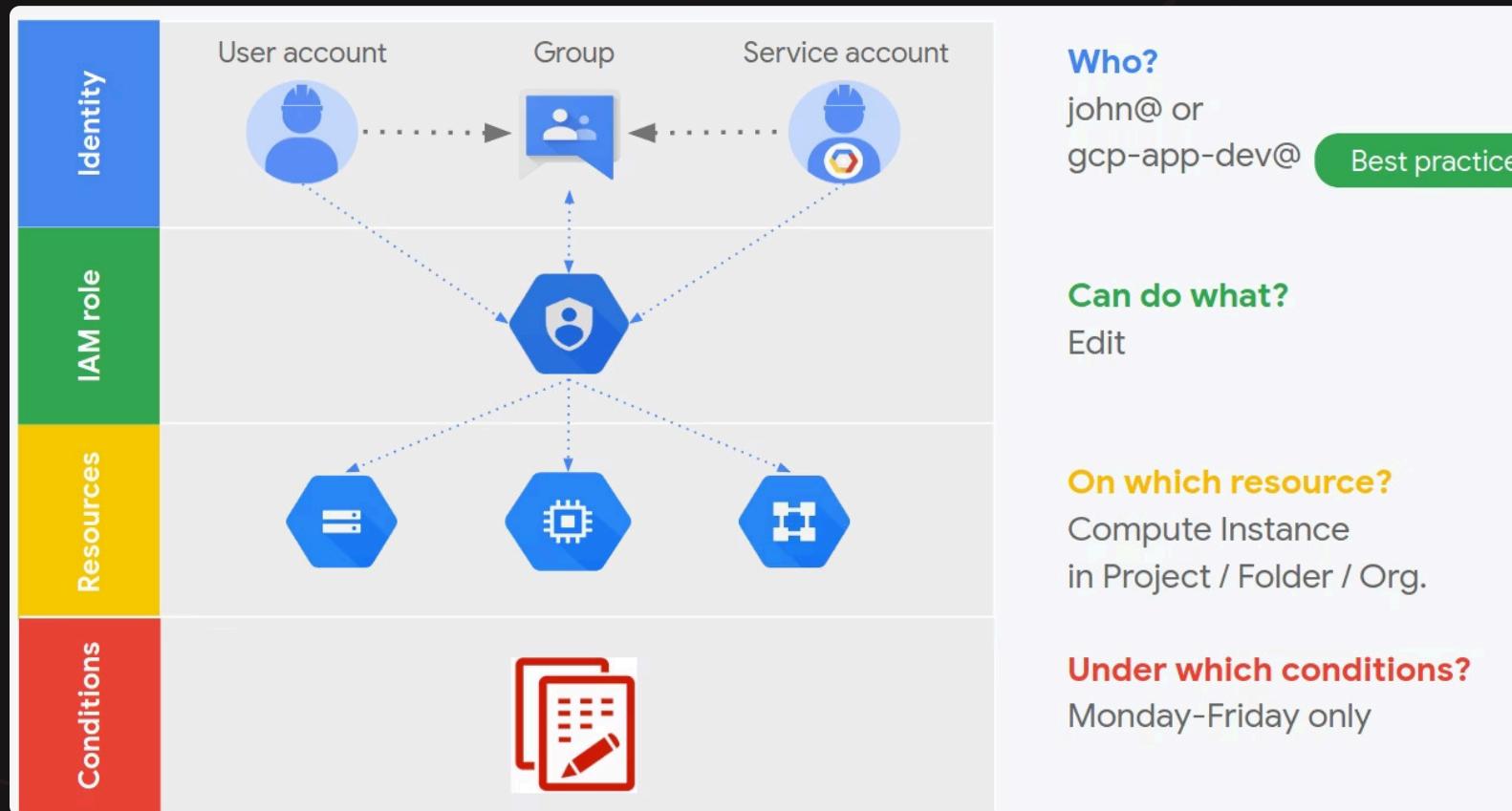


Run

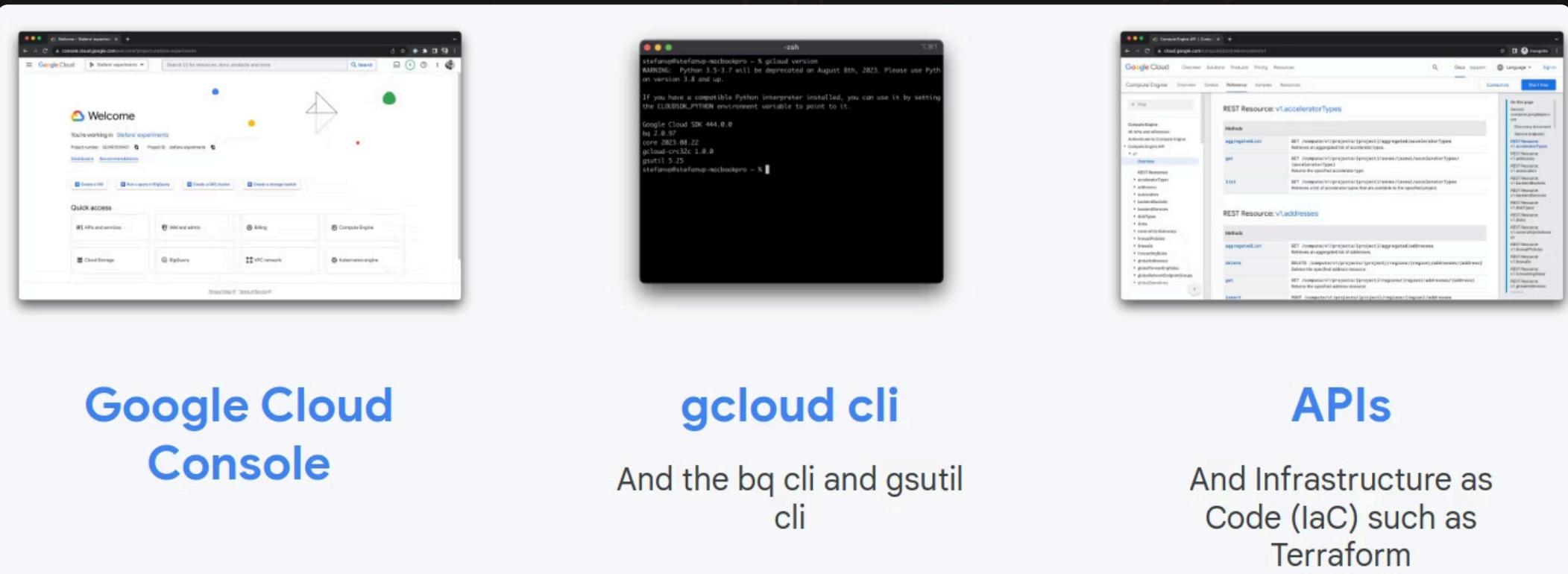
Stuff you want to do VS stuff you have to do



Cloud IAM policy



How do I interact with the cloud?



Google Cloud Console

gcloud cli

And the bq cli and gsutil
cli

APIs

And Infrastructure as Code (IaC) such as Terraform

A word of warning

Working in the cloud is...hard.

- ⚠ Everything taking longer because extra layer of communication
- ⚠ Syntax can be hard to remember
- ⚠ A lot of services to be confused about
- ⚠ Permissions, permissions,

The only way to learn is to use it. If you can, start simple and then scale up

Lets take a look

<https://console.cloud.google.com>

The screenshot shows the Google Cloud Platform dashboard for the project "My First Project". The left sidebar lists pinned services: Billing, IAM & Admin, APIs & Services, Marketplace, Compute Engine, Cloud Storage, VPC network, Cloud Run, SQL, Kubernetes Engine, and BigQuery. The main content area has tabs for DASHBOARD, ACTIVITY, and RECOMMENDATIONS. The DASHBOARD tab is selected, displaying sections for Project info, APIs, Google Cloud Platform status, Monitoring, and Error Reporting.

Project info

- Project name: My First Project
- Project number: 1063505572188
- Project ID: promising-flash-335109

API APIs

Requests (requests/sec)

Time	Requests/sec
9:45	1.0
10 AM	0.8
10:15	0.6
10:30	0.4

No data is available for the selected time frame.

Google Cloud Platform status

Google Cloud SQL
Global: Issues with Cloud SQL for MySQL instance migration to 5.7 when source databases have gtid_mode set to ON.
Began at 2021-12-07 (12:44:21)
All times are US/Pacific
Data provided by status.cloud.google.com

Monitoring

- Create my dashboard
- Set up alerting policies
- Create uptime checks
- View all dashboards

Error Reporting

No sign of any errors. Have you set up Error Reporting?

News

Meme of the day

**Me signing up for my 4th GCP account
to get the \$300 of free credits**

