GOOGLE CLOUD MEETUPS

HAPPTIQ × POSEDIO

Agenda

Date: 04.07.2022 Time: 17:30 Location: Talent Garden

17:45 - 18:00	**Introduction**
18:00 - 18:45	**Topic #1 + Quiz:** *From sensor to dashboard* - Don't give up and don't close your eyes - How easy it can be to overcome big data challenges and visualize results in Dashboards for various stakeholders with the Google Cloud, **Muris Kavlak**
18:45 - 19:00	**Break & Networking**
19:00 - 19:45	**Topic #2** *Out of the Dark - Into the Light - durchblickers' story of renewing its infrastructure from multi-cloud VM based services to mainly Cloud Run based services in the GCP*, **Michael Karner**
19:45 - 20:30	**Networking**

HAPPTIQ X POSEDIO

HAPPTIQ

Your reliable and competent partner in
innovative cloud solutions and integrations

Born in the Cloud

- Founded in March 2020
- Google Cloud Partner since Mai 2020
- Focus
  - GCP Application- and Infrastructure Modernization
  - Google Workspace
- Advisor and Foundational work

- 19 Google Cloud Professional Certs
- 4 Google Cloud Authorized Trainer
- 50 Customers in DACH
- 15 Google Cloud Expertises
- 5 Public Customer Success Stories

*“Our **mission** is to enable companies to focus on their core business by leveraging cloud native technologies.”*

*“Our **vision** is a world where IT just works.”*

Your Speakers

Muris Kavlak

Co Founder, happtiq

Gökhan Sevilmiş

Co Founder, happtiq

Your Speakers

Muris Kavlak

Co Founder, happtiq

Why ?

Google: “Every company - regardless of size or industry - will differentiate itself from its competitors through technology. Largely, that technology will be in the form of software. Great software is centered on data. Thus, every company is or will become a data company.”

- Our world is moving faster
  - We need real-time instead of batch
  - Data size is increasing
- We want new use-cases
  - Machine learning
- Huge knowledge gaps

The Goal

- Build a reliable solution for
  - Ingesting huge amounts of data
    - Enable IoT solutions
  - Storing and analyzing datasets
  - Visualizing results
- Have minimal operational overhead
- Possibility for additional use-cases
- Overcome common challenges

Challenges with Big Data

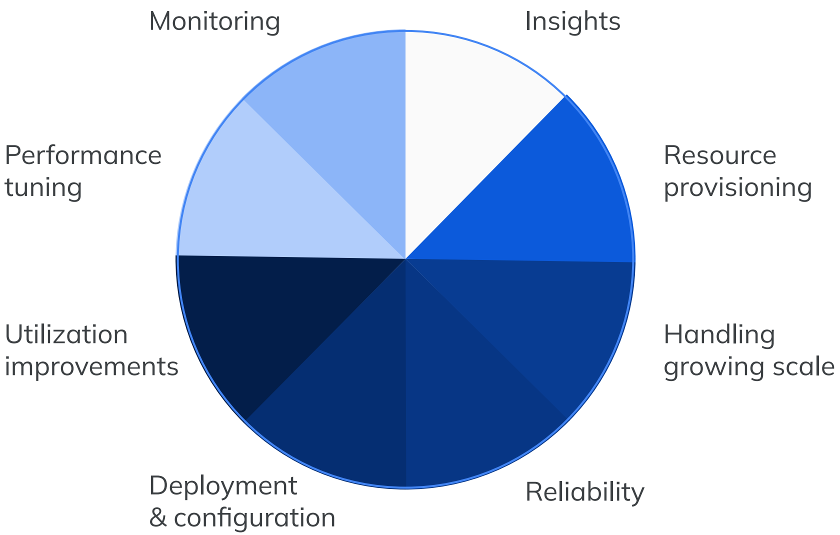
Variety

Volume

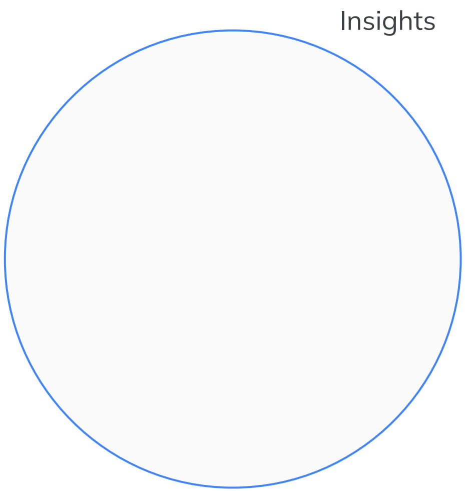
Velocity

Veracity

Let Google do the heavy work



Typical Big Data Processing



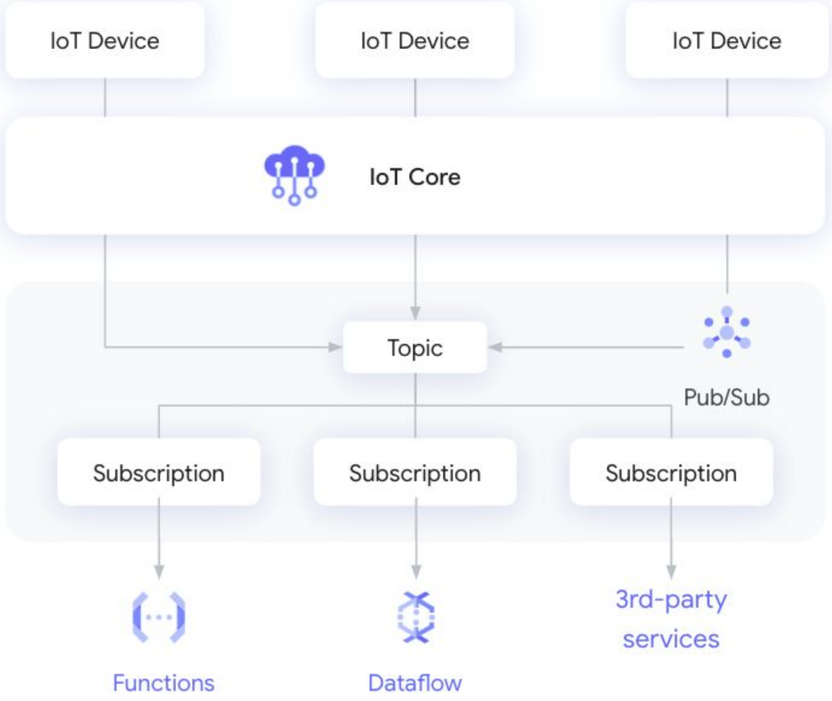
With (Google) Cloud

Build a real-time data solution with Google Cloud

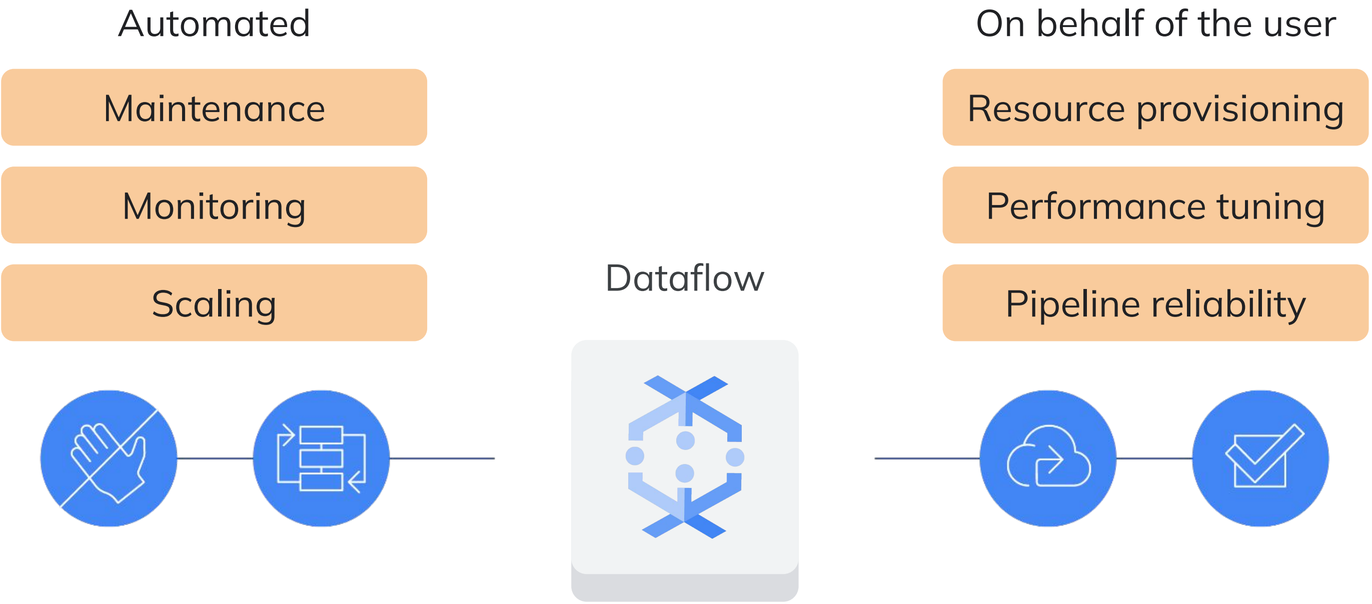
Pub/Sub	Dataflow	BigQuery	Data Studio	Looker
Ingest streaming data	Process streaming data	Store and analyze	Visualize the results	

Ingest with Pub/Sub

- Scalable messaging service
- Durable message persistence
- Global service by default
- Real-time features with Dataflow & Cloud Functions
- Loose coupling of application & devices



Process with Dataflow



Automated

Maintenance

Monitoring

Scaling

Dataflow

On behalf of the user

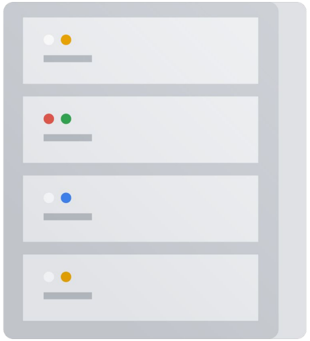
Resource provisioning

Performance tuning

Pipeline reliability

Store & analyze with BigQuery

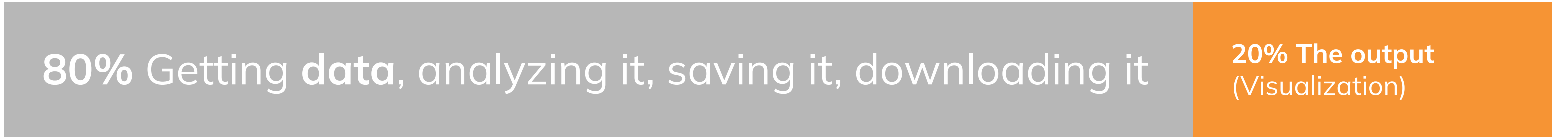
- Load batch & stream data
- Completely serverless
- Query PB of data ad-hoc with SQL
  - + federated queries
- Built-in Machine learning
- Integrate with various visualization tools



Data warehouse

The 80/20 Rule

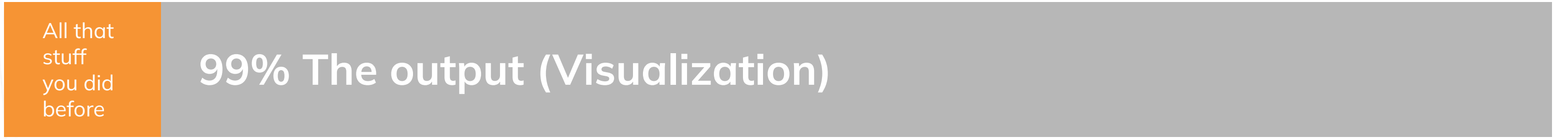
What you spend your time on:



80% Getting **data**, analyzing it, saving it, downloading it

20% The output
(Visualization)

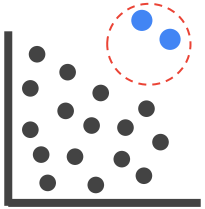
What your audience actually cares about:



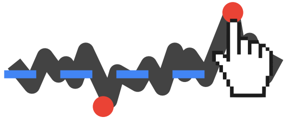
All that
stuff
you did
before

99% The output (Visualization)

Use Visualization



Visualizing a dataset allows you to spot hidden trends



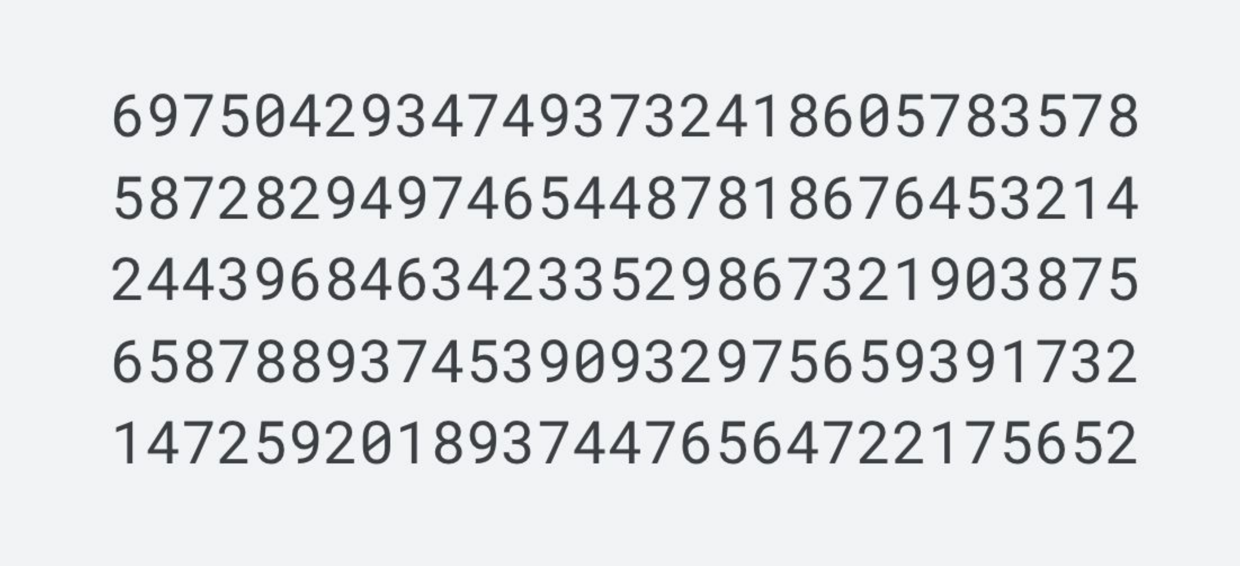
Interacting with a dataset visually is often faster than writing SQL

ACME
53,206
66,104
327,396
47.29%
How are site sessions trending?
What are the top countries by sessions?
Which channels are driving engagement?

Deliver powerful insights to your audience through reports

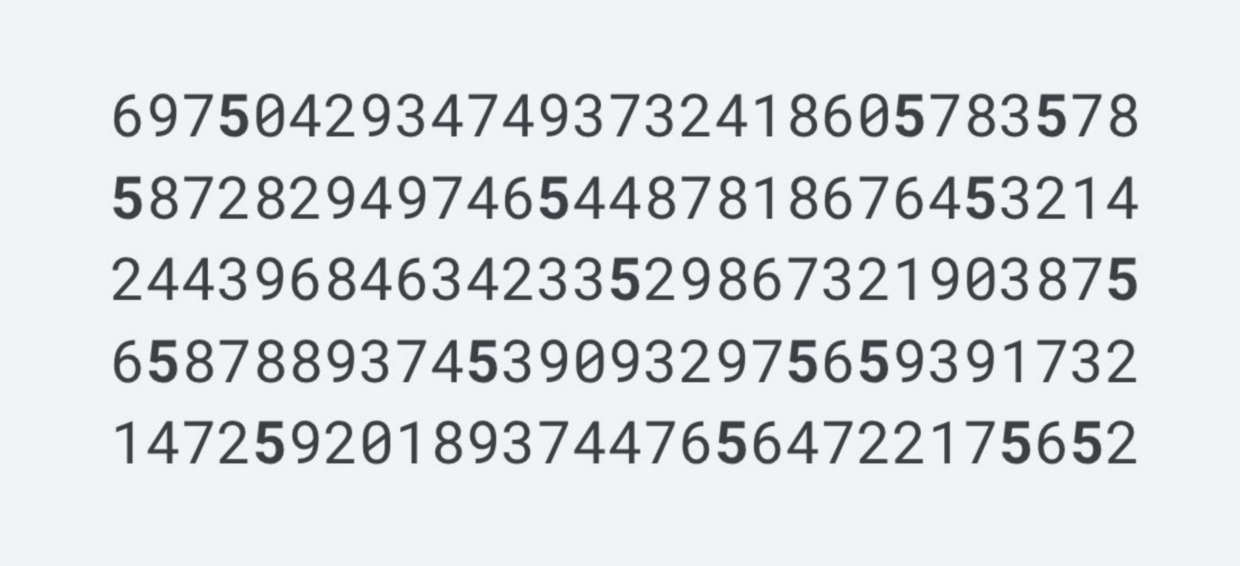
Get scalable performance as your dataset grows with BigQuery-backed visualization tools

Visualization Theory - Count the fives



```
69750429347493732418605783578
58728294974654487818676453214
24439684634233529867321903875
65878893745390932975659391732
14725920189374476564722175652
```

Visualization Theory - Count the fives



697**5**04293474937324186O**5**783**5**78

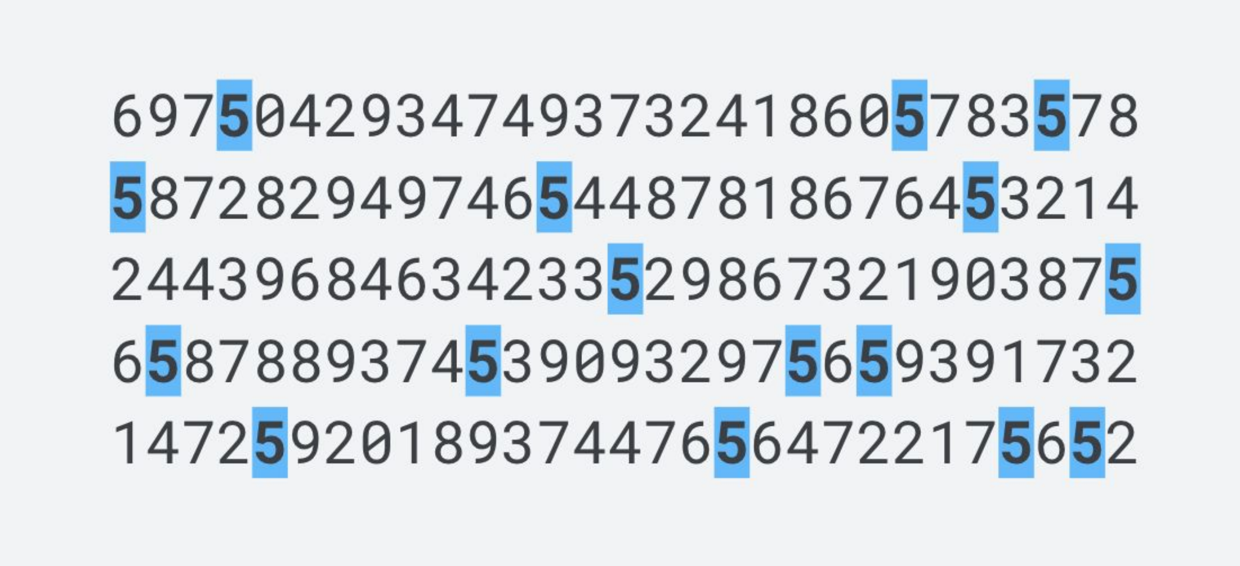
587282949746**5**44878186764**5**3214

24439684634233**5**2986732190387**5**

6**5**87889374**5**39093297**5**6**5**9391732

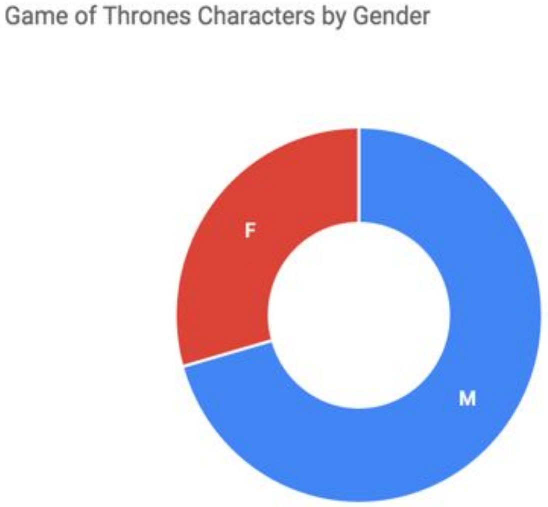
1472**5**920189374476**5**6472217**5**6**5**2

Visualization Theory - Count the fives

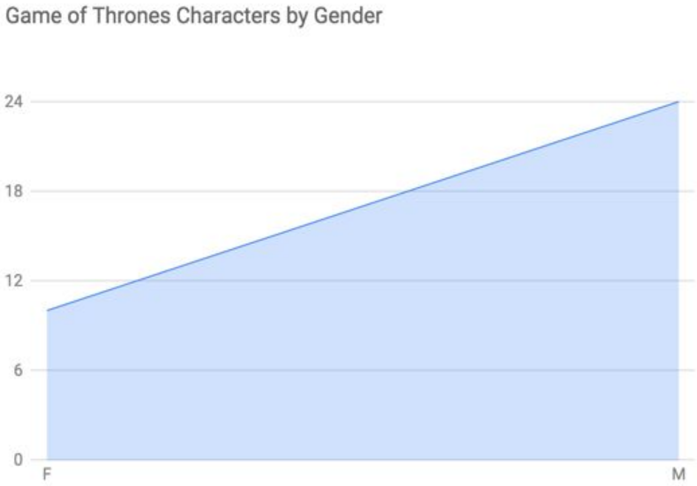


```
69750429347493732418605783578
58728294974654487818676453214
24439684634233529867321903875
65878893745390932975659391732
14725920189374476564722175652
```

You choose - effective or ineffective



OR



Live Demo

Q&A

Quiz

HAPPTIQ X POSEDIO

GOOGLE CLOUD MEETUPS

HAPPTIQ × POSEDIO

durchblicker

Out of the Dark - Into the Light (?)

durchblickers' story of renewing its infrastructure
from multi-cloud VM based services
to mainly Cloud Run based services in the GCP

durchblicker

The company

- ➔ Austria’s independent and market-leading online price comparison website
- ➔ Roughly over 80 employees
- ➔ 28 price comparisons (for insurance, energy, financial products, telecommunications)
- ➔ Founded in 2010 by Michael Doberer and Reinhold Baudisch and was sold to Netrisk Group earlier this year
- ➔ Netrisk group operates leading Internet portals for online comparisons and for taking out insurance as well as other fix cost offers in CEE

About me

- ➔ Software developer since end of the 90’s
- ➔ Working for durchblicker since 2019 as Senior Software Engineer
- ➔ Started in backend - nowadays mainly focused in the infrastructure/operations/on-call area

What is this talk about?

- ➔ Short version: Why and How we migrated our infrastructure into GCP
- ➔ Disclaimer: Retrospective view - nobody should be blamed

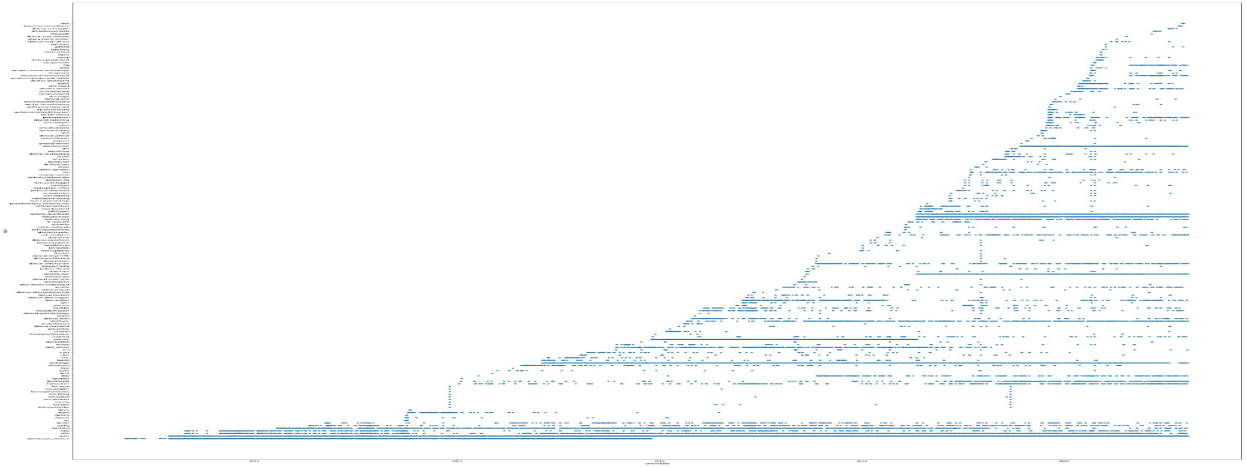
From 2010 (durchblicker founded) until now

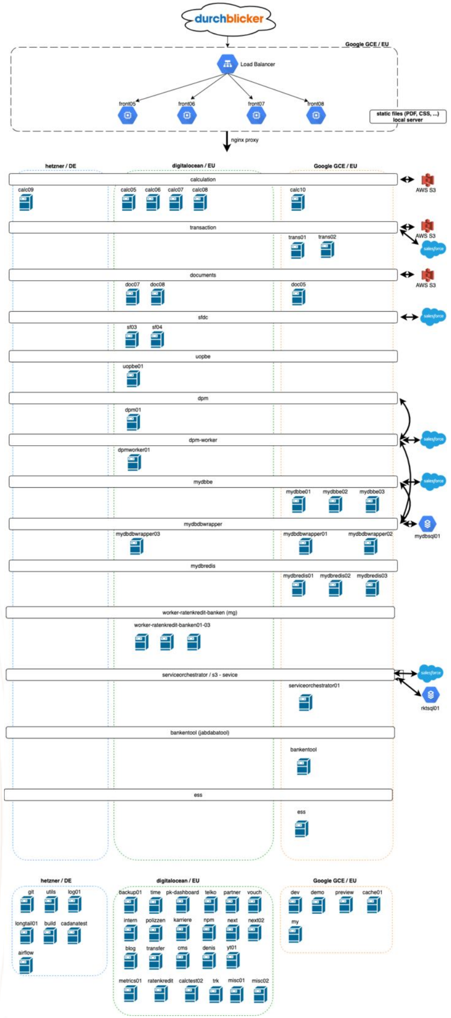
- ➔ 2010: Ancient (dark) times with bare metal servers, Virtual Machines and jQuery - first release of git was in 2005
- ➔ docker - *2013
- ➔ Kubernetes - *2014
- ➔ AWS Lambda - *2014 // Google Cloud Functions - *2016
- ➔ AWS Fargate - *2017 // Google Cloud Run - *2019
- ➔ GKE Autopilot - *Feb 2021

- ➔ “Startup feeling” - trial and error of ideas
- ➔ “Graveyard of ideas” - codebase and infrastructure

State of durchblicker in early 2021:

- ➔ Around 120 git repositories on our self hosted gitlab server (+ 40 not in use anymore) - graphical commit history





Infrastructure

- ➔ Around 60 VMs (+bare metal) mainly in GCE, digital ocean, hetzner
- ➔ DNS (route 53) - AWS
- ➔ A few S3 buckets
- ➔ MySQL DBs in GCP and locally

- + Around 20 (Core-)(Micro-)Services
- + Anxiety by a few developers

(dark part) We identified major pain points (non exhaustive list)

➔ We build too much by ourselves instead of using Saas/PaaS

➔ Secret management not handled optimally

➔ Not optimal build processes on Jenkins

➔ Logging/Alerting is not good (Graylog, Elastic, Grafana run by ourselves)

➔ Too much noise due to a large number of non-errors

➔ VMs need constant manpower to maintain (security patches, OS Updates, …)

➔ Configuration of services in production not optimal

➔ Slow setup of new test systems

➔ Lost trust in Ansible repo as single source of truth regarding state of the servers

➔ Certificate management

➔ Multi cloud issues (network traffic)

➔ No living DevOps culture

➔ Security

A few assumptions were taken

- ➔ We have to fully migrate all our core services into one cloud
- ➔ Want to make best-practice decisions - need help of consultants
- ➔ We have to containerize the services
- ➔ No big bang - we have to migrate the services one by another
- ➔ We will take a rather conservative approach - we migrate the services as they are (as possible). Our hypothesis was that we can distinguish between bugs already existing in the old world and bugs due to migration
- ➔ Central secret management is a must have
- ➔ Infrastructure as code is a must have - no more manual changes or mouse clicks in the infrastructure
- ➔ 3 stages: development, staging and production
- ➔ Good logging / Error reporting
- ➔ Replace Jenkins by Gitlab

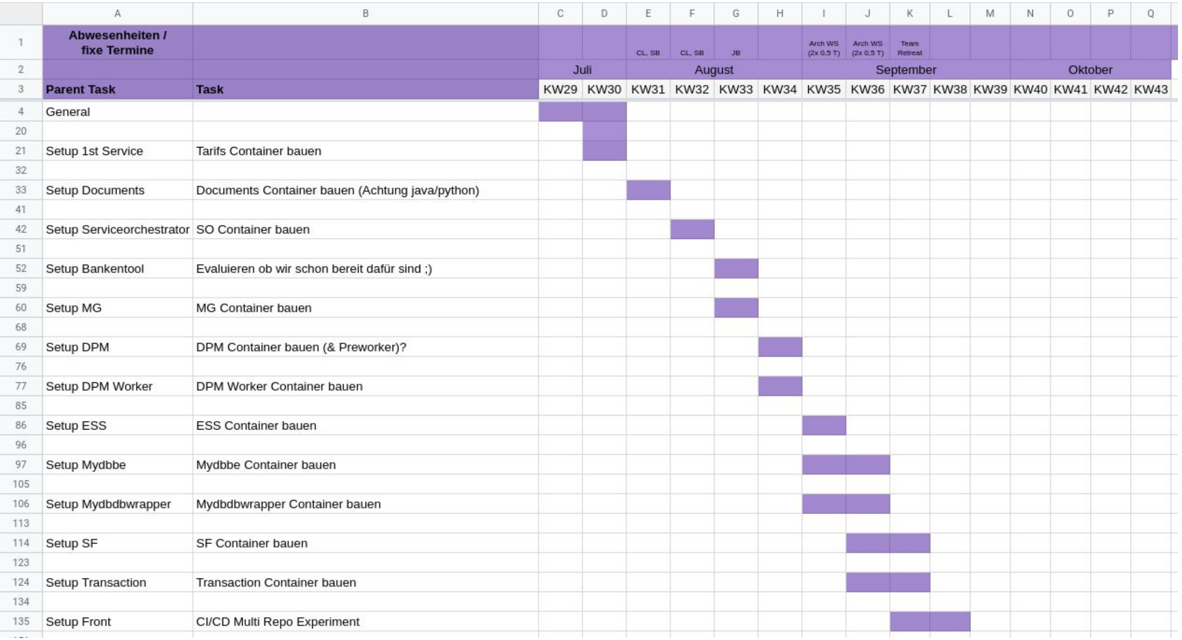
durchblicker

A decision was made

- ➔ PoC migration of one service comparing AWS and GCP with support
  - ◆ AWS - Consultant
  - ◆ GCP - Google
- ➔ Outcomes:
  - ◆ CDK (Cloud Development Kit) vs Terraform
  - ◆ Count of products in the cloud
  - ◆ Logging / Searching in the logs
  - ◆ UI/UX
- ➔ No clear winner
- ➔ 11-June-2021: More or less a gut decision for GCP
  - ◆ Imperative vs declarative definition of the infrastructure
  - ◆ GCP worked well for us
  - ◆ Better support for Kubernetes (if needed)
- ➔ Searching for consultant - we chose happtiq
- ➔ 05-July-2021: Kickoff workshop with happtiq and Google and the journey began

The beginning - migration roadmap - a complete underestimation

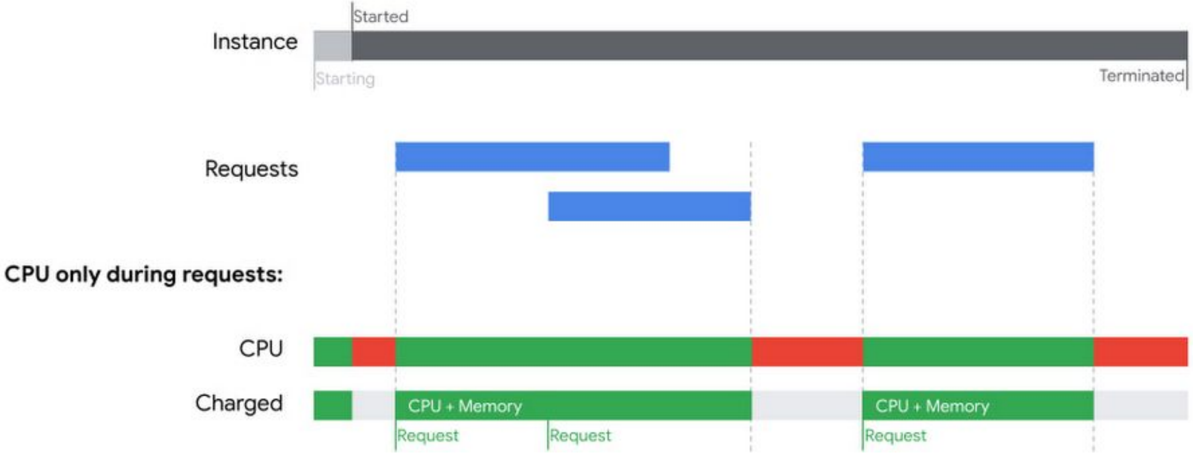
- ➔ Learning phase
- ➔ Migrate 1-2 services per week into Cloud Run
- ➔ Done until October 2021



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	**Abwesenheiten / fixe Termine**				CL, SB	CL, SB	JB		Arch WS (2x 0,5 T)	Arch WS (2x 0,5 T)	Team Retreat						
2			Juli		August				September					Oktober			
3	**Parent Task**	**Task**	KW29	KW30	KW31	KW32	KW33	KW34	KW35	KW36	KW37	KW38	KW39	KW40	KW41	KW42	KW43
4	General		■	■													
20				■													
21	Setup 1st Service	Tarifs Container bauen		■													
32																	
33	Setup Documents	Documents Container bauen (Achtung java/python)			■												
41																	
42	Setup Serviceorchestrator	SO Container bauen				■											
51																	
52	Setup Bankentool	Evaluieren ob wir schon bereit dafür sind ;)					■										
59																	
60	Setup MG	MG Container bauen					■										
68																	
69	Setup DPM	DPM Container bauen (& Preworker)?						■									
76																	
77	Setup DPM Worker	DPM Worker Container bauen						■									
85																	
86	Setup ESS	ESS Container bauen							■								
96																	
97	Setup Mydbbe	Mydbbe Container bauen							■	■							
105																	
106	Setup Mydbdbwrapper	Mydbdbwrapper Container bauen							■	■							
113																	
114	Setup SF	SF Container bauen								■	■						
123																	
124	Setup Transaction	Transaction Container bauen								■	■						
134																	
135	Setup Front	CI/CD Multi Repo Experiment									■	■					

tarifs - first learnings

- ➔ tariff - calculation engine (oldest service about 11 years old)
- ➔ Only a few dependencies and a lot of traffic to let us see failures quickly
- ➔ August 2021: We were done with containerization and terraform stuff. The Cloud Run service was running for a few hours. Woohoo???
- ➔ The first issue - RabbitMQ and CPU throttling of Cloud Run



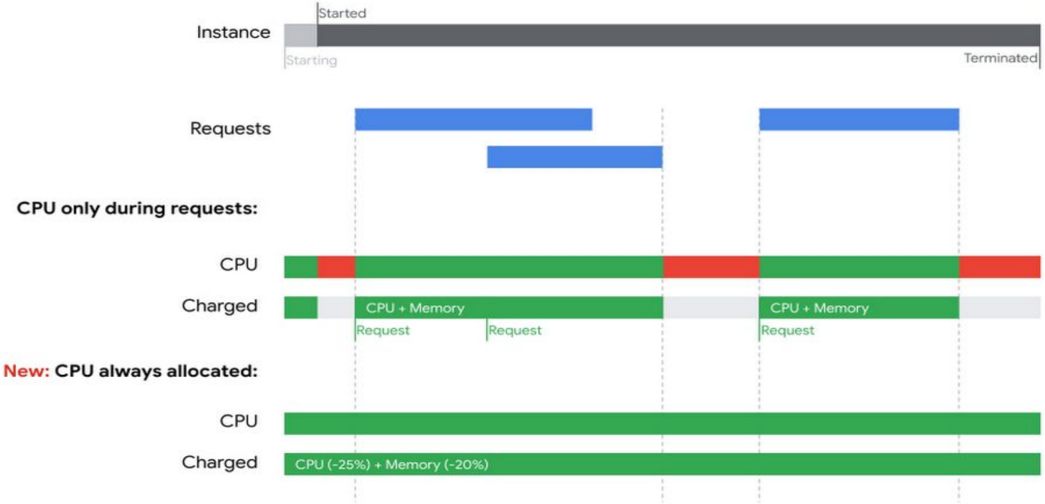
happtiq - we have a problem

- 2 solutions:
  - Replace RabbitMQ by PubSub and use Push via HTTP
  - Use GKE instead of Cloud Run for all services using RabbitMQ
- tarifs in GKE/Autopilot
  - We set up everything in terraform to run the tarifs Service in a GKE/Autopilog cluster
  - Problem solved - but: Load test showed - scale up is too slow
  - GKE without Autopilot? Never! We want as much managed as possible.
  - Splitting? Run one instance in Autopilot and the others in Cloud Run?

What should we do? Some kind of **miracle** would be nice :-)

The miracle happened

“CPU always on” feature in Sept 2021”
→We migrated 3 services from Autopilot to CR, 1 was already running in CR
→ We gained speed



durchblicker

By the end of 2021 - we:

- ➔ used terraform to organize the org and 3 projects for the stages in GCP
- ➔ set up IAM
- ➔ set up org policies
- ➔ set up billing alerts
- ➔ set up alerting in Atlassian OpsGenie
- ➔ migrated more than the half of our production services
- ➔ migrated a few MySQL databases
- ➔ introduced a few Cloud Functions (helper functions or nanoservices)
- ➔ were using Cloud Storage
- ➔ were using Cloud Scheduler (cron equivalent)
- ➔ were using Secret Manager in GCP
- ➔ were using Cloud Armor to limit access to some resources
- ➔ were using Cloud Debugger/Cloud Profiler
- ➔ extended the GCP bunyan node logger to our needs (Error reporting)
- ➔ had a k8s/Autopilot cluster for the gitlab runners created (balloon pods)
- ➔ got to know kaniko (build container images in a k8s cluster)
- ➔ have our own container registry
- ➔ still had fun to work on the project

durchblicker

The adventure of migrating the frontend (since Jan 2022)

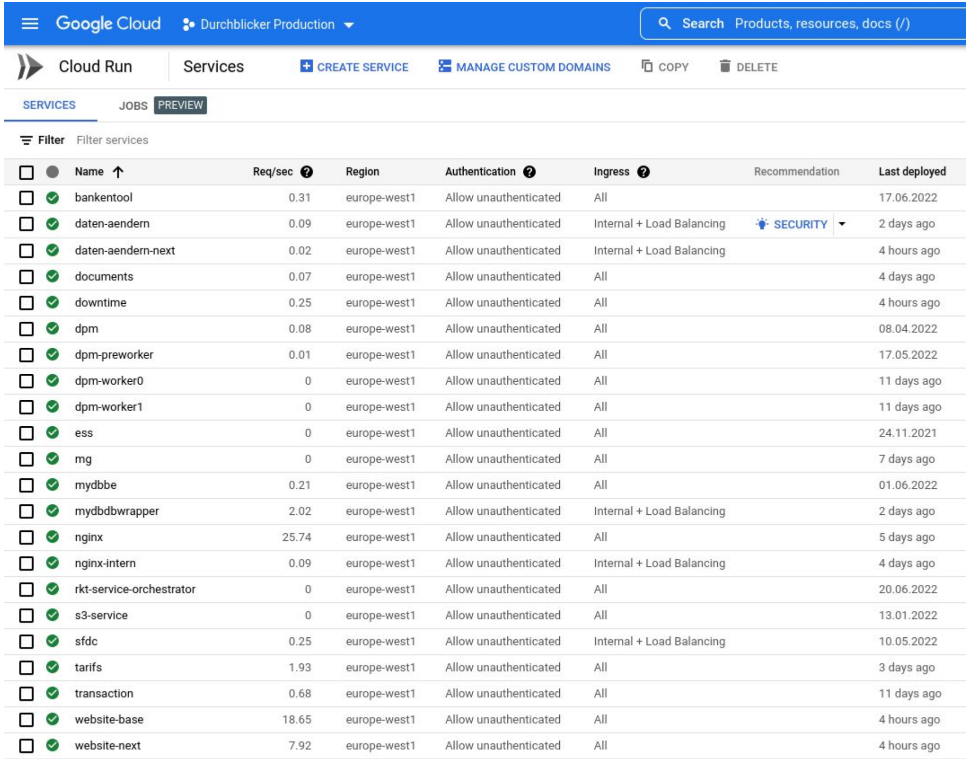
- ➔ Consisting of 11 git repositories
  - ◆ Old style (legacy) website
  - ◆ New nextJS/React based website
  - ◆ Configuration data
  - ◆ Service for publication/removal of PDF documents in downloads
  - ◆ JSONs (>100k Files), binaries (sum up to around 9 GB)
- ➔ Built on Jenkins - which glued together the repos, built the artifacts, rsynced/pushed it to the 4 front servers.
- ➔ Frontservers are GCE based VMs, each having each own reverse proxy nginx behind Load Balancer of GCP
- ➔ Monster? - Definitely yes ;-)

A few problems we faced:

- ➔ CI/CD pipelines in gitlab. The repos have a topological ordering. (trigger builds in other repos)
- ➔ Upload to Cloud Storage is rather slow if you have a large number of files (run the gitlab jobs with 10 cpus)
- ➔ Caching in gitlab jobs is slow (we use image layer cache of docker)
- ➔ Load Balancer - URL Maps - Path Rules: Path rules do not use regular expressions or substring matching (needed a nginx Cloud Run service)
- ➔ Release: We have a deployment monolith. We need to release 4 services in Cloud Run and copy data to 5 buckets nearly at the same time.

Since 01-June-2022 we run in production with the website and we are still alive :)

And here we are now



□	●	Name ↑	Req/sec	Region	Authentication	Ingress	Recommendation	Last deployed
□	✓	bankentool	0.31	europe-west1	Allow unauthenticated	All		17.06.2022
□	✓	daten-aendern	0.09	europe-west1	Allow unauthenticated	Internal + Load Balancing	SECURITY	2 days ago
□	✓	daten-aendern-next	0.02	europe-west1	Allow unauthenticated	Internal + Load Balancing		4 hours ago
□	✓	documents	0.07	europe-west1	Allow unauthenticated	All		4 days ago
□	✓	downtime	0.25	europe-west1	Allow unauthenticated	All		4 hours ago
□	✓	dpm	0.08	europe-west1	Allow unauthenticated	All		08.04.2022
□	✓	dpm-preworker	0.01	europe-west1	Allow unauthenticated	All		17.05.2022
□	✓	dpm-worker0	0	europe-west1	Allow unauthenticated	All		11 days ago
□	✓	dpm-worker1	0	europe-west1	Allow unauthenticated	All		11 days ago
□	✓	ess	0	europe-west1	Allow unauthenticated	All		24.11.2021
□	✓	mg	0	europe-west1	Allow unauthenticated	All		7 days ago
□	✓	mydbbe	0.21	europe-west1	Allow unauthenticated	All		01.06.2022
□	✓	mydbdbwrapper	2.02	europe-west1	Allow unauthenticated	Internal + Load Balancing		2 days ago
□	✓	nginx	25.74	europe-west1	Allow unauthenticated	All		5 days ago
□	✓	nginx-intern	0.09	europe-west1	Allow unauthenticated	Internal + Load Balancing		4 days ago
□	✓	rkt-service-orchestrator	0	europe-west1	Allow unauthenticated	All		20.06.2022
□	✓	s3-service	0	europe-west1	Allow unauthenticated	All		13.01.2022
□	✓	sfdc	0.25	europe-west1	Allow unauthenticated	Internal + Load Balancing		10.05.2022
□	✓	tarifs	1.93	europe-west1	Allow unauthenticated	All		3 days ago
□	✓	transaction	0.68	europe-west1	Allow unauthenticated	All		11 days ago
□	✓	website-base	18.65	europe-west1	Allow unauthenticated	All		4 hours ago
□	✓	website-next	7.92	europe-west1	Allow unauthenticated	All		4 hours ago

Sum-up

We have reached many of our goals. And there’s still a lot to do.

Cons:

➔ Google support is kind of adventurous - one example

- Issue was reported on 14-Jan-2022
- 23-Feb-2022: “Thank you for your patience. I have received an update from the Cloud Run Product team and unfortunately we are unable to find the root cause of the issue due to unavailability of the logs for that time frame. However If you face the similar issue again, I would request you to please inform us immediately on this case within 30 days even after this case is closed.
  However, If you face the issue again even after 30 days, you can create a new case referring to this case number and we will help you on the issue.

  Thank you for your understanding. Please let me know if you have any further questions.”

➔ AWS has more modern features than GCP

➔ A few limitations are not understandable

➔ The cloud console app on smartphones is weird (is this really from Google?)

durchblicker

Pros:

- ➔ Terraform is pretty cool
- ➔ Working with happtiq is great
- ➔ GCP is stable
- ➔ Cloud Run is a really cool product - I am curious what the future brings
- ➔ My fellow colleagues (the gang of four)

And -

We are hiring!!!!

durchblicker

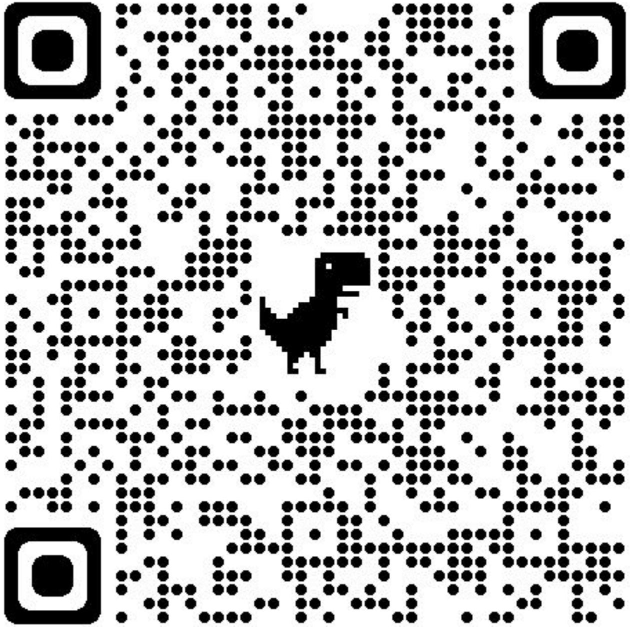
Thank you and time for questions

SAVE THE DATE!

Summer Event - August 2022

HAPPTIQ X POSEDIO

It’s Feedback time!



HAPPTIQ X POSEDIO