

GOTO AARHUS 2021



About Me

Mathias Schwarz

PhD in CS from Aarhus University

Senior Software Engineer II at Uber

Love to build infrastructure software

6 years into building stateless platforms at Uber



Software Engineering at Planet Scale

June 11, 2021

Mathias Schwarz, Senior Software Engineer II

Uber

The Uber Infrastructure team @aarhus





- Microservice deployment
- **Runtime Configuration**

Scale: Business

18M

Trips happening every day, based on Q1 2020

Uber Uber Freight Uber JUMP Eats **Uber Elevate** POSTMATES Cornershop

Scale: Infrastructure

4000

Services across several of our own data centers plus cloud zones such as AWS



Scale: Deployment

58K

Builds / week

Production deploys / week

5K



Stateless services at Uber



Hybrid cloud

More elastic (if you are small)

Special purpose hardware



Cloud Zones / POPs

AZs with virtual Hardware

Cheaper

╋

Few dependencies



Private / On Prem

Core AZs with ODM Hardware

Regions and availability zones



Early days: Unstructured deploys



Important deploy system features

Consistent builds

Zero downtime

Outage prevention

Building at scale



2224 ALANALANALANALANALANALANA

Code repository

Java^{*} - GO + V

Go & Java

The preferred backend language



Monorepo management

https://eng.uber.com/go-monorepo-bazel/

Build a docker image

GHolite





Builds docker images

https://github.com/uber/makisu/ https://eng.uber.com/makisu/

Publish docker image



Upload

Makisu +uBuild

Kraken

Distributed P2P Docker Registry

<u>https://github.com/uber/kraken</u>, <u>https://eng.uber.com/introducing-kraken/</u>

Structured deploys with µDeploy

Uber

UBER ENGINEERING'S MICRO DEPLOY: DEPLOYING DAILY WITH CONFIDENCE



In 2014, Uber began expanding ever rapidly. Our platform grew from about 60 cities to 100 in the spring, and then to 200 in the fall. Meanwhile, our fastest growing cities were among our oldest.

As the number of additional platform engineers grow, so did the disorganization of deploying new code. Each team used its own custom shell scripts to shepherd new versions of its microservices into production, manually monitoring them with service-specific tools. When upgrading hosts went awy, engineers tediously rolled back one machine at a time. With more and more engineers working on Uber services, this manual labor couldn't scale and sometimes prolonged outlages.

How did we learn to consistently deploy every day? We developed Micro Deploy (known as µDeploy for short), our in-house deployment system that builds, upgrades, and rolls back services at Uber.

The Daily Deployment Process

Uber engineers use Micro Deploy once their code is production-ready—that is, once it's reviewed, accepted, passing all unit tests, and merged into the repository. First, the engineer selects a service to upgrade in the µDeploy interface. To start an upgrade workflow, they select a deployment narr effer to a version of the source code in the Git repository.

Rollout to clusters, not servers



Scheduler+Manager for compute clusters

https://github.com/uber/peloton

https://eng.uber.com/resource-scheduler-cluster-management-peloton/

Mesos Logo, by The Apache Software Foundation, Apache License, v2.0

Deploy into cluster in zones



Safety: Monitoring metrics

Execution Plan		← → c	https://umonitor.uberinternal.com/alerts/	b068ef75-3f47-48d0-9e18-0628feb693;	36	
		III Apps	🗲 Compute team 🧔 Overseer 🕚 Clusto Web	o 🏟 🗲 Hailstorm 🗋 driving cost per m	👏 mesos master 🌓 Aurora master 👘 localhost 🕻	l clusto changelog 🔳 uMonitor API 🚦
pelaton job publican production canary upgrading slack notified slack channel publican events umonitor Monitoring 15 alerts	time_done		ERT CROUPS > UCS: CATEWAY > UNEQUAL O Unequal object count betw of services/getconfig_gateway teamous STATUS PASS Time Servies (1)	BJJECT COUNT BETWEEN ZONES URRENT VALUE (M3) 2	CHITTCAL THRESHOLD 2	WARN THRESHOLD 1.5
PHX			mv produktion name catelo eljecti do runtime, em v produktion service delest config- gateway type gaige (maximum) subtratt env grodskuton service object config- gateway type gaige (minimum)			
Continuous monitorir performance	ng of business and e metrics		Object ID count Flipr Events	•	Wear 07-49 AM	Wed IDES AM
			U	Monit	or	

time-series based rollback triggers

Safety: Whitebox integration



Explicit part of the rollout plan

Hailstorm

Load tests Whitebox integration tests

Safety: Continuous blackbox



Blackbox testing/monitoring

Virtual trips in production

Efficiency at scale

Multi Cloud

Fully managed

Predictable

Alternatives

"Deploy across multiple cloud providers including AWS EC2, Kubernetes, Google Compute Engine, Google Kubernetes Engine, Google App Engine, Microsoft Azure, Openstack, Cloud Foundry, and Oracle Cloud Infrastructure, with DC/OS coming soon"



https://spinnaker.io/

Efficient infrastructure with Up



Deploy into regions



- → C (a) up.uberinternal.com/p	project/tax-calculations#publican				* • • •) 🔓 O 🗯 🌒 i
tax-calculations	RODUCTION PROD STAGING REPLAY	Actions ~ uOwn tax-calculations				
publican			TIER 2 LIFE CYCLE STADE active	✓ Deploy	phabricator/base/14644111-2-9¢95867b8 do compute-balancer 18 Jul, 19:30 CEST	
 phabricator/base/14780787 phabricator/base/14644111- 	-1-g741c38fd H-XS Feature XYZ -2-gc95667b8 H-XS Patch XYZ-1			✓ Deploy	phabricator/base/14644111-2-gc95667b8 compute-balancer 18 Jul, 13:30 CEST	
CANARY			16 INSTANCES RUNNING 16 INSTANCES REQUESTED	✓ Deploy	phabricator/base/14644111-2-gc95667b8	
DCA			426 INSTANCES REQUESTED		17 Jul, TROUCEST	more details
PHX			610 INSTANCES RUNNING 610 INSTANCES REQUESTED			
· · · · ·			ELK Log Explorer Statistics			
T Deploy phabri T Deploy phabri kri 21 Jul,	cator/base/1478#787-1-g741c38fd istiyan 11:30 CEST	OPERATION IS PAUSED	ð Abort			
Executio	on Plan					
DCA	peoton. Peloton job subican production canary upgrading Back notified slack channel publican events					
	umonitor Monitoring 15 alerts	tme done 🗸				
PHX						
	- COLLAPSE EXECUTION PLAN	VIEW ACTION DETAILS ->				٢

Up - tax-calculations - product X +		
← → C ■ up.uberinternal.com/project/tax-calculations#publican		🖈 🚨 🕲 😘 😋 🌸 🌒 :
tax-calculations	Actions ~ uOwn tax-calculations	
p		16 INSTANCES RUNNING
CANARY		16 INSTANCES REQUESTED
		126 INSTANCES DUNNING
DCA		
PHX		610 INSTANCES RUNNING
T		610 INSTANCES REQUESTED
		ELK Log Explorer Statistics
L	ELK Log Explorer Statistics	
Deploy phabricator/base/14780787-1-g741c38fd	OPERATION IS PAUSED	
21 Jul, 11:30 CEST	🛧 Continue 😏 Rollback 🕲 Abort	
Execution Plan		
peloton, Peloton job publican production canery upgrading		
slack notified slack channel publican events	V	
uncenter Mentering 16 state	time_done	
	\checkmark	
PHX		
- COLLAPSE EXECUTION PLAN	VIEW ACTION DETAILS -+	

Up - tax-calculations - product X + C wuberinternal.com/project/tax-calculations#publican		
tax-calculations PRODUCTION INFOS		
publican	TICK 2 Deploy Deploy Deploy Deploy Control States (States States	2-yc35657b8
Plan	ته على ال-على (1-عاد 201) الأسلي (1-201) من المحمد الم	2-yc35667120
peloton Peloton job publican.production.canary upgrading		
slack notified slack channel publican events	time done	
umonitor Monitoring 15 alerts	, 	~
PHX		

● ● ○ Up - tax-calculations - produ: X + → C ● up.uberinternal.com/project/tax-calculations#publican		x 🖸 🛛 🖓 😘 🖉 🛪 🌒 :
tax-calculations PRODUCTION Prop STAGING REPLAY Actions V John tax-calculations		
publican	TER 2 phabrics.tor/base/166 LIFE CYCLE STAGE active & Deploy	14111-2-gc3566788
Executio	- Depley	4111-2-0(5566756
peloton Peloton job publican.production.canary upgrading		
slack notified slack channel publican events	time done	
umonitor Monitoring 15 alerts	✓	~
PHX		

tax-calculations PRODUCTION FROM	ations		
publican	TIER 2 STAGE active	oloy Daabricator/Base/1664111-2-qc3553783	
e pharistotor/henr/latin.	√ Det	-16 JUL 19 JUL (251)	
peloton Peloton job publican.production.canary upgrading		-	
slack notified slack channel publican events			
umonitor Monitoring 15 alerts.			
PHX	~		
	TION DETAILS -		

Challenges in abstraction



Avoiding manual migrations



Plan for migrations

- More productive teams
- More reliable / available
- Faster
- Lower cost of capacity

- Cloud <-> On-prem zones
- Zones come and go
- Cluster rolls change
- New cluster technologies
- New container runtimes
- New base images

Goals

implies

A continuous need for improving efficiency and reducing the cost of running the company

Infra changes

Must not have a negative impact on the goals

Let's add a new zone



Let's add a new zone



Declarative Configuration

Horizontal Service: public	Scale an Environment: production	×
Environ Recomr Scaling	nent is not scaled according to recommendations. nendations calculated 2 days ago. Metrics (> Autoscaler Documentation (>	
Region	Scale to	
DCA	250 instances Autoconfer 0 Recommendation: 183 There are 5 canary instances in region DCA	1500
РНХ	250 instances	1500
Canary Turning on C	anary will place 5 canary instances in DCA. Canary instances will be deployed before other instances.	
Autoscaling Autos	caling is disabled	
Standard	y •	•
		Submit





Declarative configurations

- Region placement
- Dependencies

Continuous evaluation

Optimal application of the configuration

← → C (▲ up tax-calcula	suberinternal.com/project/tax-calculations#publican	Actions -			* 🖸 🚱 🚱 🚱 🏠 🦄 :
publica		Comm dar Uncomora	TIER 2 LIFE CYCLE STADE RO	Deploy DTT	
phabric	ator/base/14780767-1-g741C3816 (I-X3 Feature XYZ ator/base/1464111-2-g55667b8 (I-X3 Patch XYZ-1		~	Deploy	phabricator/base/14644111-2-gc9566
CANAR	·		16 INSTANCES RUNNING 16 INSTANCES REQUESTER 426 INSTANCES RUNNING		18 Jul, 13:30 CEST
PHX			426 INSTANCES REQUEST 610 INSTANCES REQUESTED 610 INSTANCES REQUESTED ELK Log Explorer Statistics		
11 04	phaper catory/basey/84768767-30741c38fc Fristiyan 21 Jul, 11:30 CEST	operation is paused * Continue う Rollback @	Abort		_
	Execution Plan DCA Pelitölin Pelittölin Pelittölin Pelittölin Pelittölin Pelittöl				_
	Idade inotified duck channel publican evens	time done	-		_
	PHX - COLLAPSE EXECUTION PLAN	VIEW ACTION DETAILS →			

Continuous improvement of large Uber services







Prometheus on M3DB

Metrics ingestion platform

Prometheus software logo, by Alexander Schwartz, Apache License, v2.0

Grafana

Graphs and dashboards

The eGraphana Logo is a trademark of Coding Instinct AB, registered in the U.S. and in other countries.

Log indexing and aggregation

Healthline



Healthline

Error log aggregation

Elasticsearch, Clickhouse, Sawmill

The elasticsearch logo is a trademark of Elasticsearch BV, registered in the U.S. and in other countries.

Log Aggregation/Queries

	Log Explorer uMonitor × +										
\leftrightarrow \rightarrow (C in umonitor.uberinternal.com/services/ch	ange-mar	nagement/logs?filters=	env%2Cp	artition%2Cdatacenter%2Cins	tance%2C%40reserve	ed.collector.filename%	62Clevel&q=lev 🕁 🤇) 📭 👘 📕 :	k =: 🧶	:
Uber	SERVICE My service		~							- (~
奋	Overview	네 Metric	s 🙆 Dashboards	≣ Tr	acing 🖬 Dependencies	⊖ Events @	Settings				
۵	Namespace Time Range										
	service:/My service 15min - now ⊻	С	Feedback Do you like	the new Lo	g Explorer experience? Yes No	Meh				💾 User Gui	de
	1 level:"error" AND env:production								0	Query Langua	.ge
ler.	Filters	+ Add	Hide Sidebar						Export Logs	ය Settings 8	3
_	env Show top 10 values	Ð			la dia seri	h-1-1-11	In the set				
0 ⁰	QQ production	2,402	0		Thu 11:39 PM		Thu 11:43 PM	Thu 11:	\$7 PM		
Ŕ	partition Show top 10 values	Û	Showing 100 results in	332 ms							
品	QQ compute-0 QQ canary	2,018 384	Timestamp 03-18 23:50:00 pm	level error	message Activity error.						
Ø	datacenter Show top 10 values	Ð	03-18 23:49:59 pm	error	Error making outbound call.						
0	QQ dcal	748	03-18 23:49:58 pm	error	Error making outbound call.						
Q	QQ dca4	673	03-18 23:49:58 pm	error	Error making outbound call.						
	QQ phx4	384 303	03-18 23:49:58 pm	error	Activity error.						
8	QQ phx2	294	03-18 23:49:57 pm	error	Error making outbound call.						
ß	instance Show top 10 values	0	03-18 23:49:57 pm	error	Error making outbound call.						
_	QQ 218103809	162	03-18 23:49:57 pm	error	Error making outbound call.						
0	QQ 33554432	156	03-18 23:49:57 pm	error	Error making outbound call.						
Q	QQ 33554434	150 147	03-18 23:49:55 pm	error	Error making outbound call.						
	QQ 16777217	143	03-18 23:49:55 pm	error	Error making outbound call.						
	QQ 16777216	140 140	03-18 23:49:55 pm	error	Error making outbound call.						
	QQ 100663296	135	03-18 23:49:54 pm	error	Error making outbound call.						
	QQ 83886081	133 127	03-18 23:49:54 pm	error	Error making outbound call.						
		-	03-18 23:49:48 pm	error	Error making outbound call.						
» _	wreserved.conector.mename Show top 10 values	U	03-10 23:49:48 pm	error	Error making outbound call.				Conditionals	a Dura	
	QQ stdout	2,402	03-18 23:49:48 DM	error	Error making outbound call.				Feedback	STR Bugs	

Distributed tracing





Jaeger

Distributed tracing for all services

https://github.com/jaegertracing/jaeger https://opentelemetry.io/

Snapshot of callgraph

All 4K microservices and how they interact

Distributed tracing

🔹 🔍 🍵 Jaeger UI 🛛 🗙 🕂					
← → C	53b604191cf1434			🖈 🖸 😘 🐑 I	* = 🌒 E
Jaeger UI Lookup by Trace ID Search	Compare System Architecture	Service Dependencies Tracing	Quality	Region	n v Help v
 hailstorm-ui: POST_/api/ getEnvironmentsForService 	a 63b604	Find		Alternate Views v	Archive Trace
Trace Start March 18 2021, 23:42:58.076 Duration 3.25s	Services 5 Depth 12 Total Spans 51	1.630		2.44e	3.250
		1.000		k. TH	0.200
<u></u>					
Service & Operation \lor > \lor »	Oms	813ms	1.63s	2.44s	3.25
hailstorm-ui POST_/api/getEnvironmentsForService					
✓ hailstorm-ui atreyu.graph.services_getenvironmentsforserv					
V hailstorm-ui atreyu.node.request					
V hailstorm-ui Services::getEnvironmentsForService					
✓ hailstorm-api Services::getEnvironmentsFor)				
✓ hailstorm-api galileo	0.03ms				
hailstorm-api uber.infra.up.api.Ser	2.14ms				
 ✓ hailstorm-api galileo 	0.02ms				
✓ hailstorm-api uber.infra.up.api.Ser	2.83s				
coconut-api uber.infra.up.api	2.82s				
coconut-api galileo	0.03ms				
coconut-api Workfl	72.88ms				
✓ coconut-api galileo	0.02ms				
coconut-api uber.inf	9.45ms				
v coconut-api galileo	0.02ms				
coconut-api uber.co	3.08ms				
v coconut-api galileo	0.02ms				
✓ coconut-api job_ins	1.82ms				
 compute-inspe 	I 1ms				
compute-in	0.74ms				
 ✓ coconut-api gallieo 	0.01ms				
coconut-api uberinf	2 728				

Summary

You can scale every aspect of software engineering

You can safely roll out 5000 times to production per week

Automation + abstractions are key to avoid constant manual migrations



Questions?





Don't forget to vote for this session in the GOTO Guide app

Backup slides

