
AWS
re:Invent
NOV. 28 – DEC. 2, 2022 | LAS VEGAS, NV

PRT279

SPONSORED BY INTEL

Mobileye: Smarter, safer autonomous solutions with Habana & Granulate

Gilberto Vargas

VP, Sales and Marketing; GM Global Datacenter Sales

Intel Corporation

Ohad Shitrit

Senior Director of AI Engineering and Algorithms

Mobileye

Pini Reisman

Director of REM Cloud Application

Mobileye

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

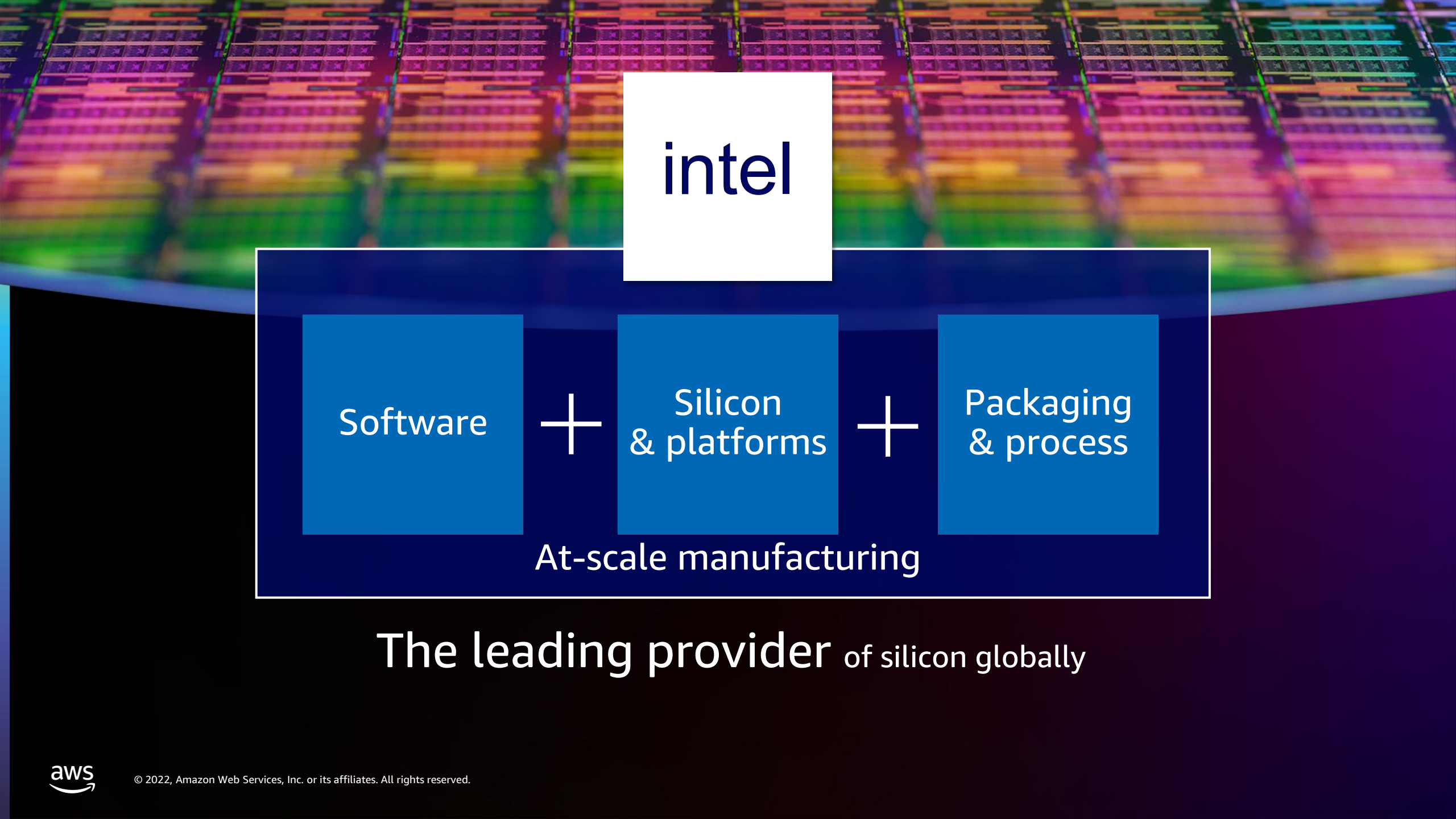


Intel and AWS

World-changing technology

A collaboration rooted in developing technology that underpins solutions to the world’s greatest challenges

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.



The leading provider of silicon globally

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Bridging physical and digital worlds with Technology superpowers

Ubiquitous =



Semiconductors are the underlying technology empowering developers and powering our customers' innovations

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

DCAI executing to accelerated growth

Growing our leadership in data center



A foundation built on Intel Xeon scalable processors

Harnessing our ubiquity and core strengths



Building on our position as a trusted industry leader

Making AI and security pervasive



Unique hardware capabilities strengthened by software

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

habana® An Intel Company | intel® | GRANULATE™ An Intel Company

Intel-powered AWS instances and software optimizations empower developers and accelerate our customers' innovations

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

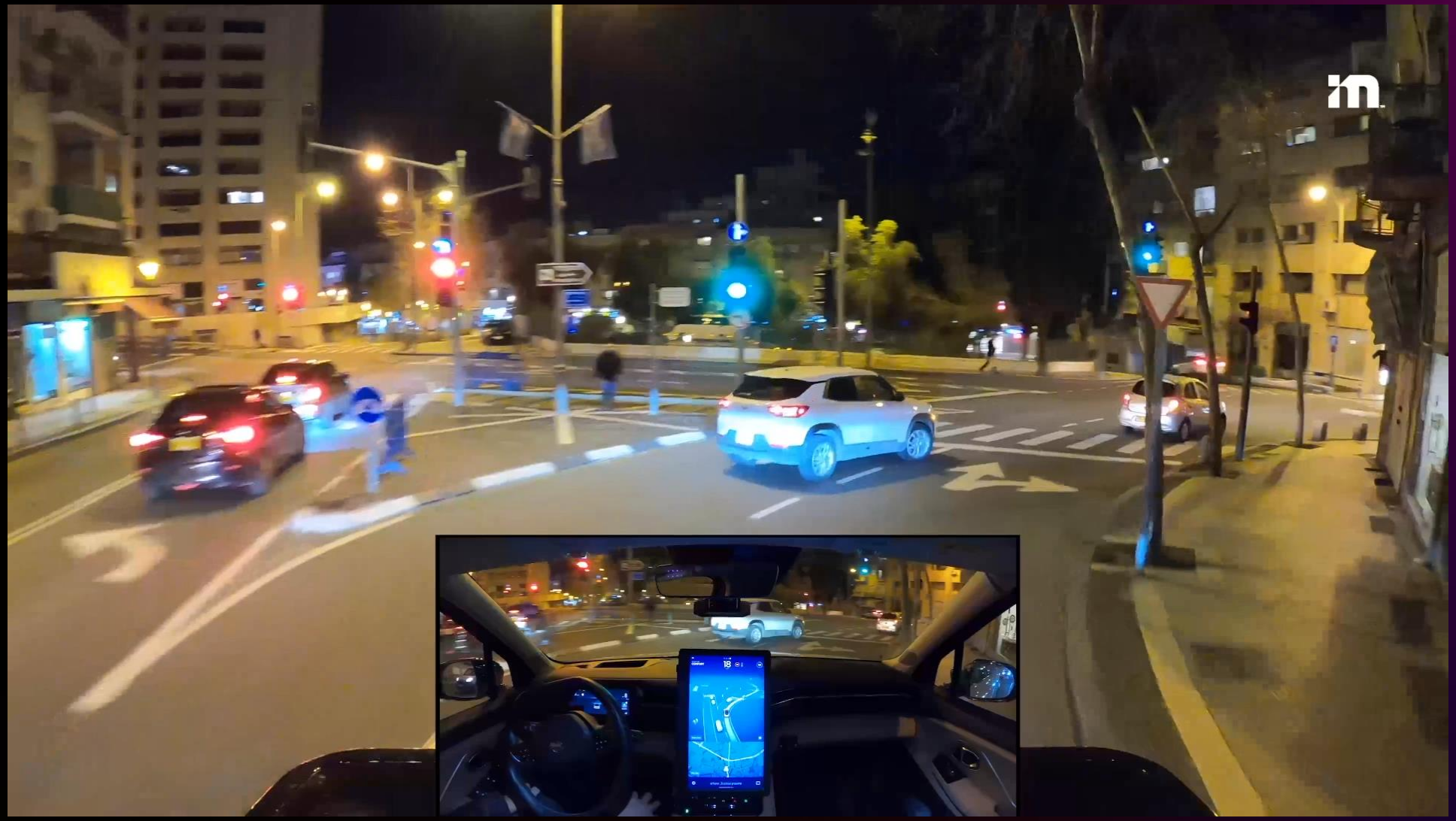
Scaling AI training at Mobileye with Amazon EC2 DL1 instances

Ohad Shitrit

Senior Director of AI Engineering and Algorithms
Mobileye

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Mobileye technologies | Autonomous driving



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Mobileye product portfolio

Driver assist

Autonomous vehicle

Base driver-assist



- EyeQ SoC
- Front camera

Cloud-enhanced driver-assist

Mobileye SuperVision



- EyeQ SoC
- 360 camera
- REM mapping
- RSS driving policy

Mobileye Chauffeur

Mobileye Drive



- EyeQ SoC
- 360 camera
- 360 Radar Lidar
- REM mapping
- RSS driving policy

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Delivering the autonomous future

Mobileye works tirelessly to deliver technology and thought leadership across the 3 pillars of autonomous driving



Sensing challenge

Harnessing the power of compute and the scale of algorithms to sense, detect, and act across independent systems



Mapping challenge

Leveraging software and hardware to hyper scale alongside the hyper growth of mapping needs across the autonomous landscape



Policy challenge

Partnering across industries governments to adopt safe and scalable policies for autonomous vehicles

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Delivering the autonomous future

Mobileye works tirelessly to deliver technology and thought leadership across the 3 pillars of autonomous driving



Sensing challenge

Harnessing the power of compute and the scale of algorithms to sense, detect, and act across independent systems



Mapping challenge

Leveraging software and hardware to hyper scale alongside the hyper growth of mapping needs across the autonomous landscape



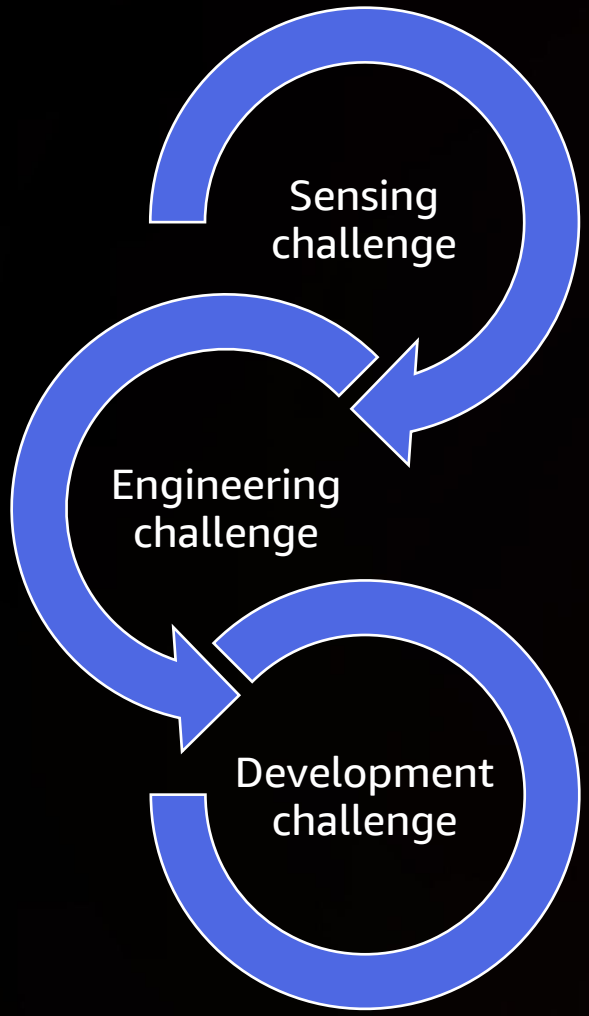
Policy challenge

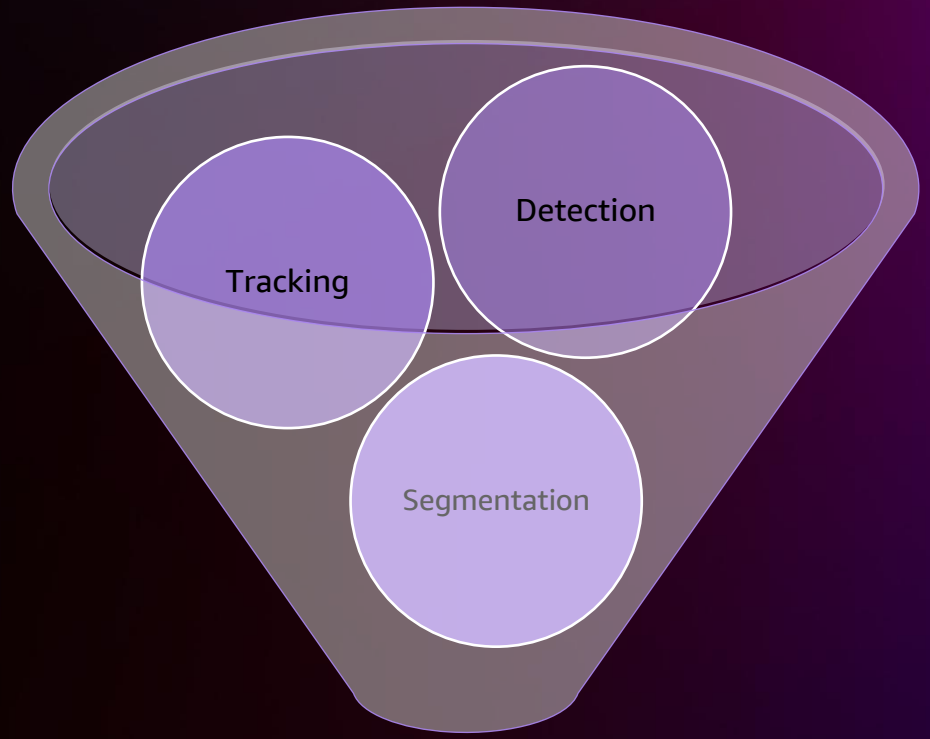
Partnering across industries governments to adopt safe and scalable policies for autonomous vehicles

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

The engineering challenge

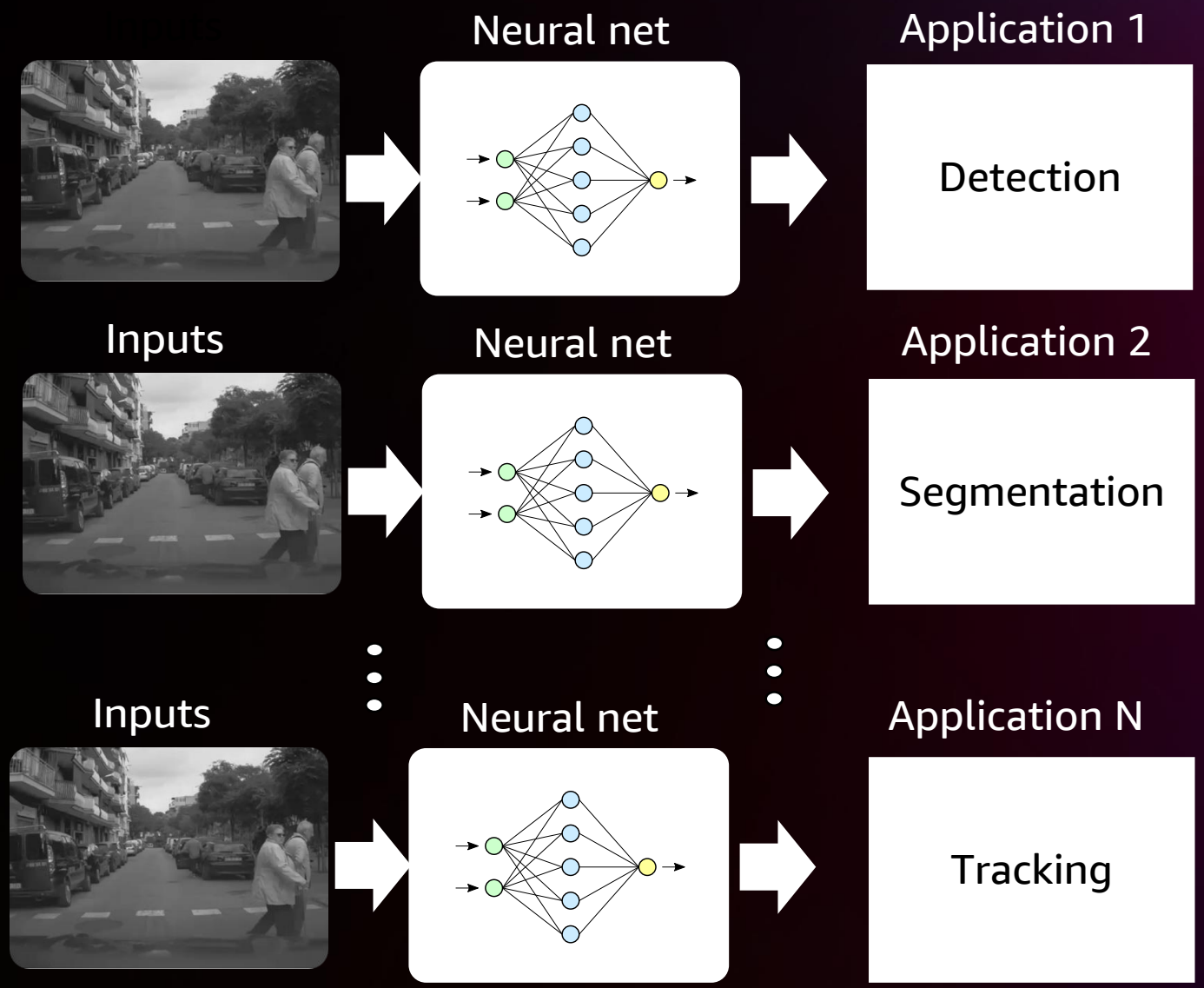






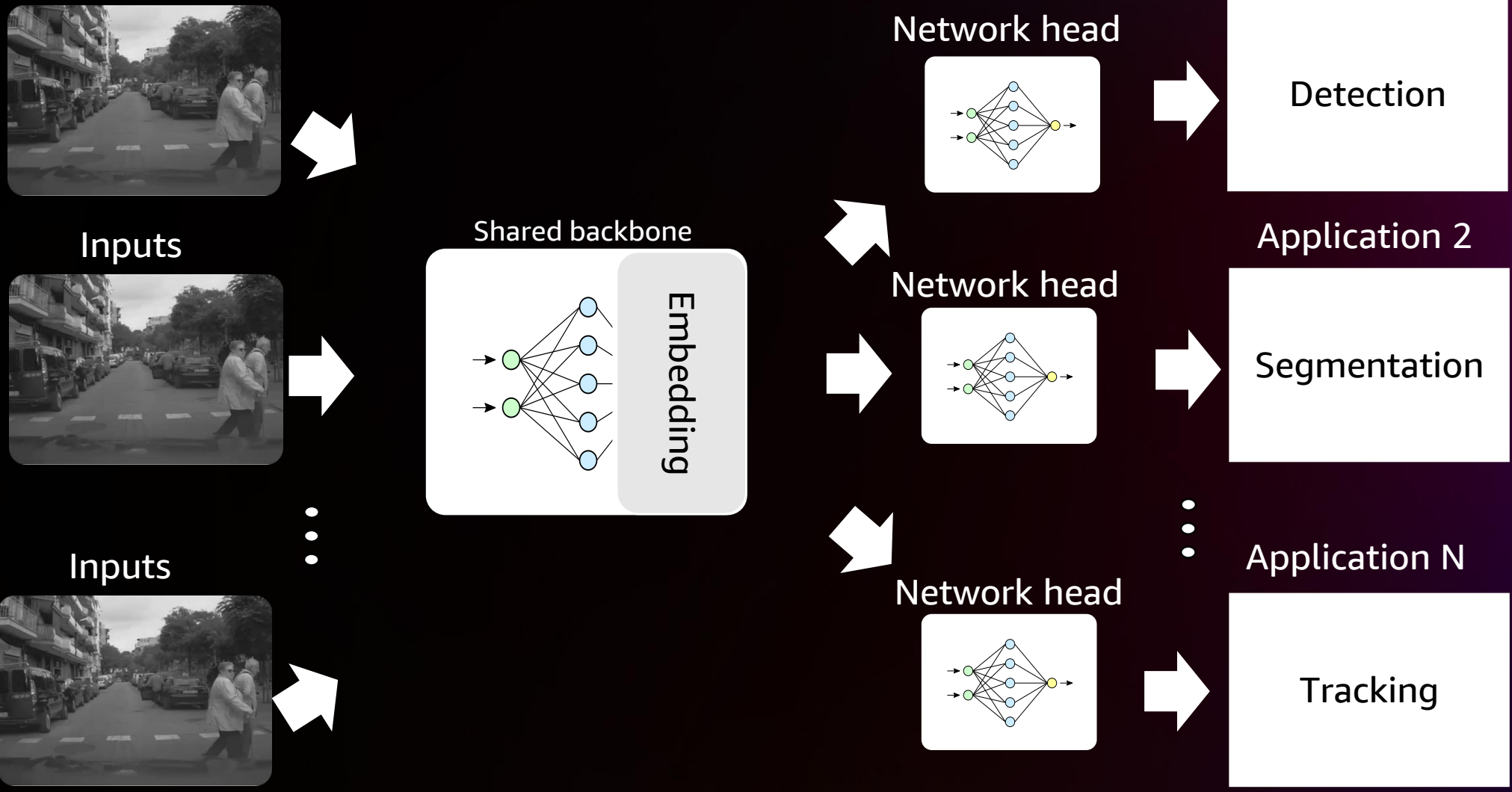
EyeQ chip

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

The solution

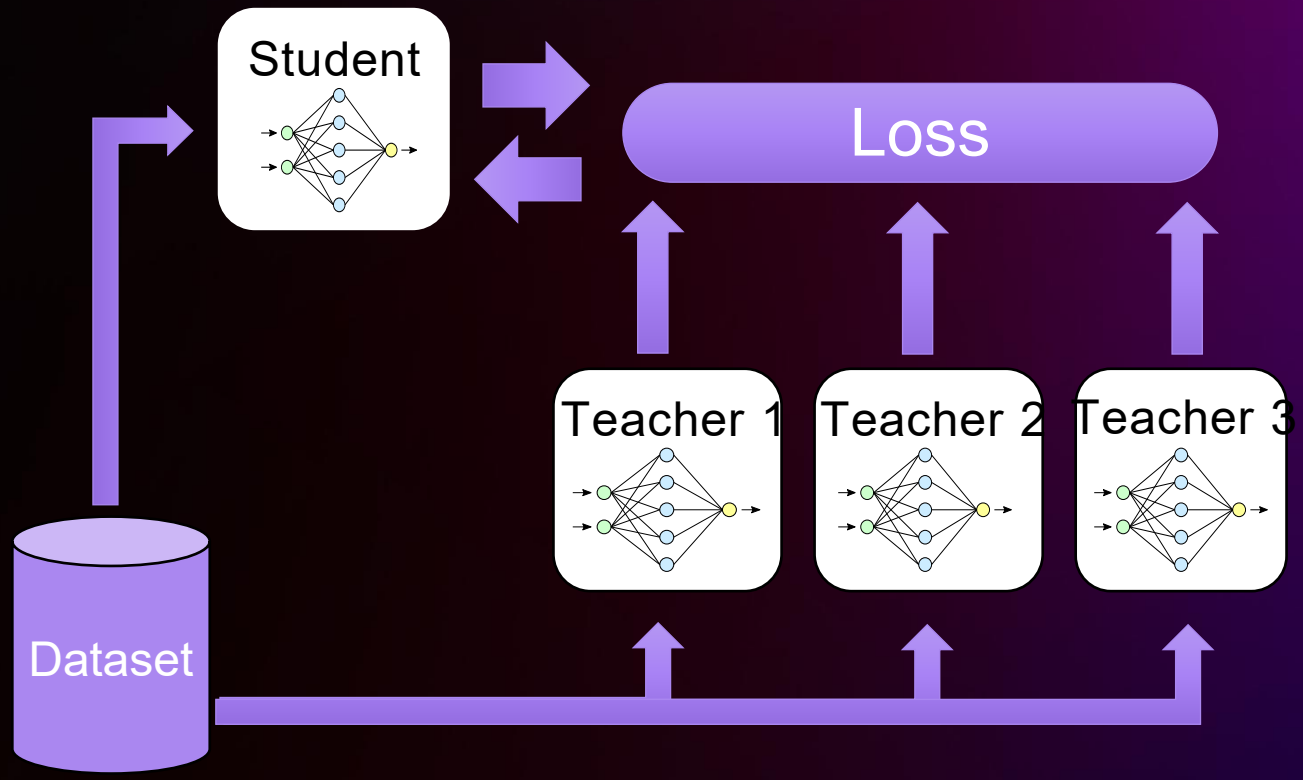


© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

The solution

Knowledge distillation

- Technologies – Teachers
- The “backbone” – Student



aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Knowledge distillation benefits



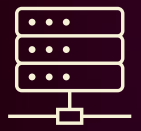
Reduce computational cost



Different technologies are decoupled



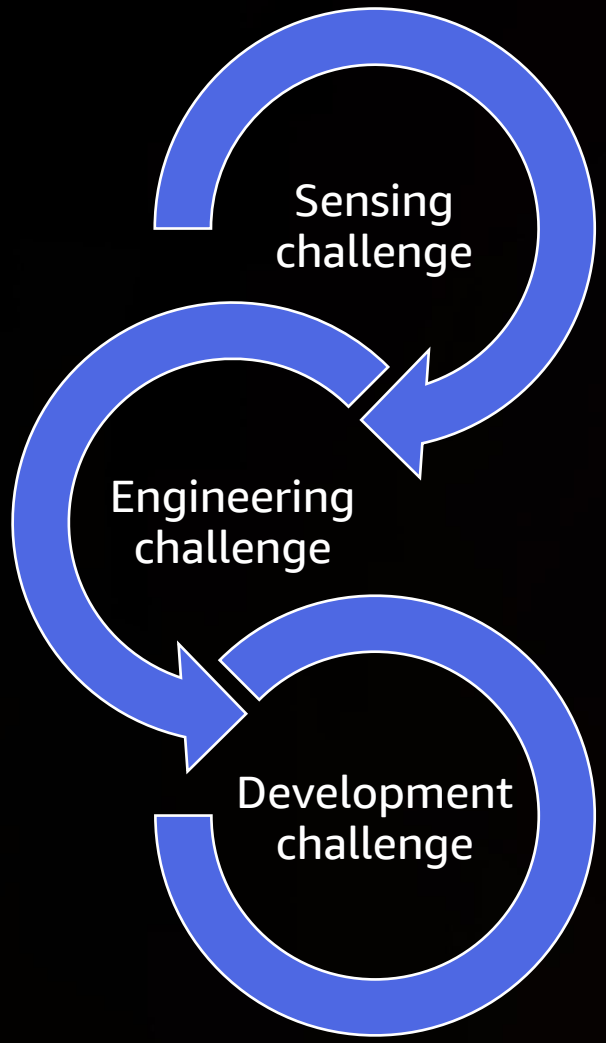
used on unlabeled data



Teachers are chip-agnostic

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

The development challenge



Problem: Mobileye's unique IP and dynamic projects require specialized but flexible compute resources. Simply put, Mobileye needs to train . . . a lot. And they require solutions that provide:

- Massive compute for research
- Short development cycles
- Flexible development environment
- Cost-effective hardware

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

The development challenge solved

To meet the needs of Mobileye's unique challenges, they turn to Habana DL1 on AWS

Habana's Gaudi instance on AWS provides

- Optimized large-scale distributed training to **accelerate the training process**
- Cost-effective, high-performance hardware that **increases the number of experiments** the team runs and ensures high accuracy



aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Experimental results

Cost-effective training

Small network

Instance	#GPUs	Batch size	Samples per second	Samples per dollar
g5.48xlarge	8	8*64	619	136,927
dl1.24xlarge	8	8*64	614	168,713
p4d.24xlarge	8	8*64	1200	131,000

23% over g5.48xlarge

28% over p4d.24xlarge

Large network

Instance	#GPUs	Batch size	Samples per second	Samples per dollar
g5.48xlarge	8	8*16	74	16,356
dl1.24xlarge	8	8*32	84	23,066
p4d.24xlarge	8	8*48	181	19,884

41% over g5.48xlarge

16% over p4d.24xlarge

aws

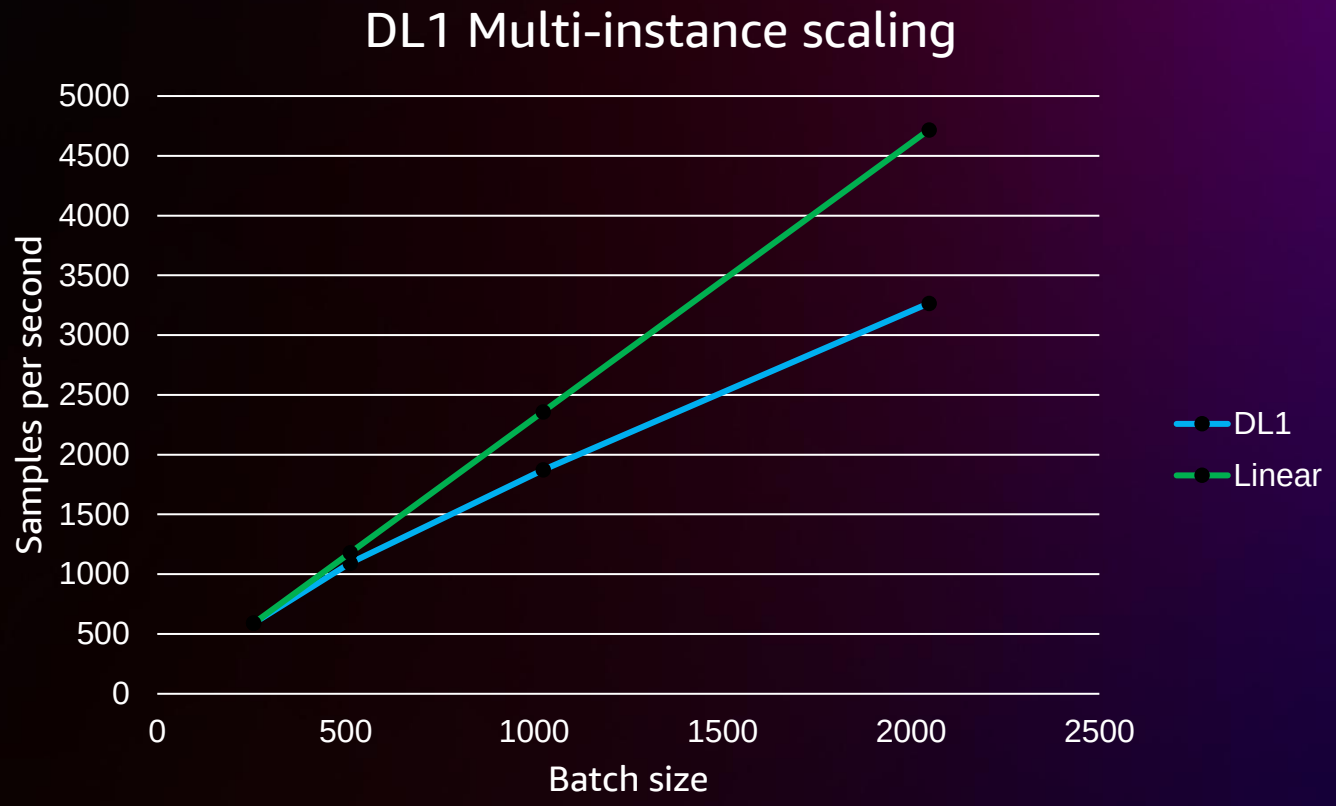
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Experimental results

Multi-instance scaling

#DL1 instances	Batch size	Samples per second
1	256	589.7
2	512	1092.0
4	1024	1872.4
8	2048	3265.6

*Expect even better performance with EFA peer-direct (coming soon)



aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Impact

Partnering with Habana is a key ingredient to driving Mobileye's mission of saving lives on the road

Cost-effective training

More experiments

Better models

Better product

Safer drive!

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

REM and Granulate

Pini Reisman

Director of REM Cloud Application

Mobileye

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Delivering the autonomous future

Mobileye works tirelessly to deliver technology and thought leadership across the 3 pillars of autonomous driving



Sensing challenge

Harnessing the power of compute and the scale of algorithms to sense, detect, and act across independent systems



Mapping challenge

Leveraging software and hardware to hyper scale alongside the hyper growth of mapping needs across the autonomous landscape



Policy challenge

Partnering across industries governments to adopt safe and scalable policies for autonomous vehicles

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

HD maps for autonomous driving

AV & L2+ systems leverage HD maps for fully & semi-autonomous driving



mapbox

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

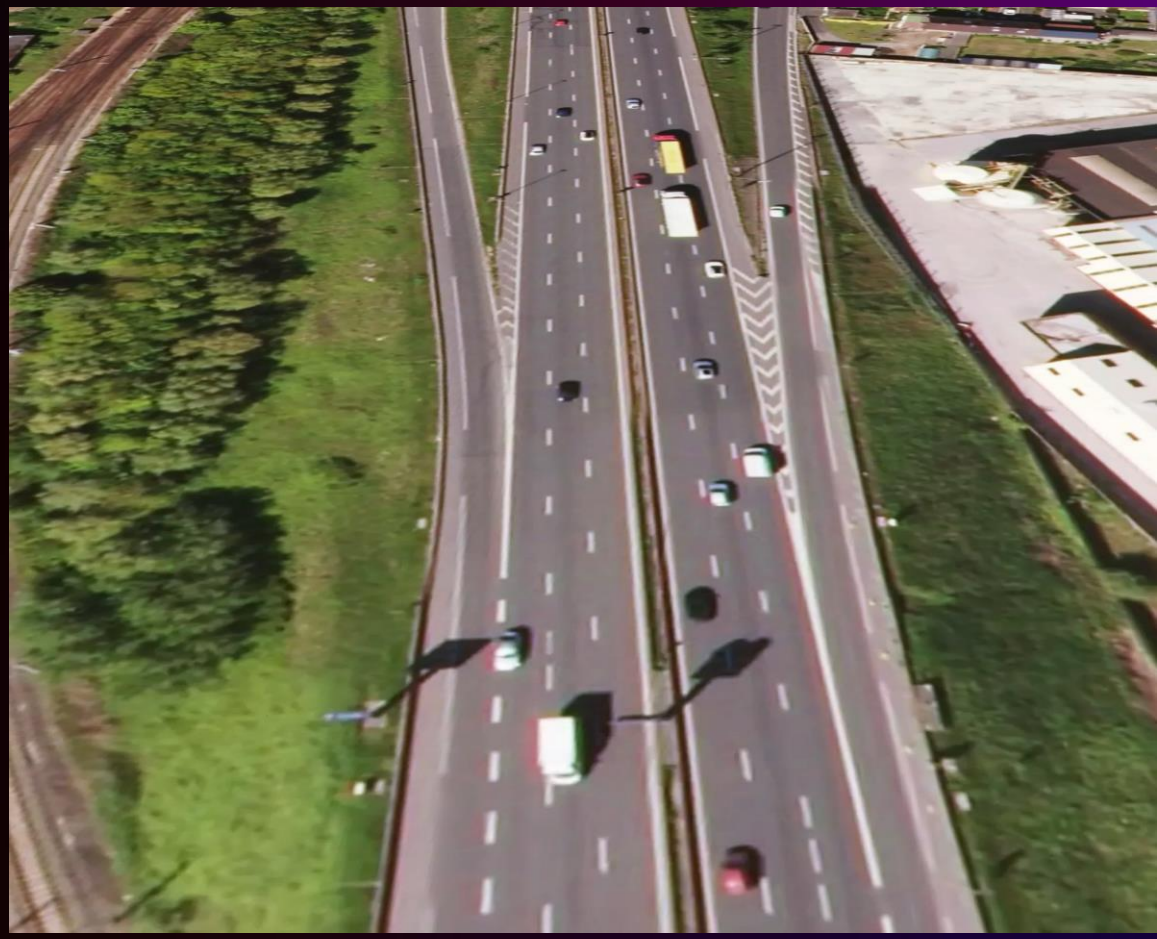
REM: Map from crowd-sourced ADAS systems

Crowd-source Mobileye consumer ADAS systems

Transform to representation of the road

- Geometry
- Driving semantics
- Traffic rules

- ➔ RoadbookTM (Mobileye's AV map)



mapbox

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

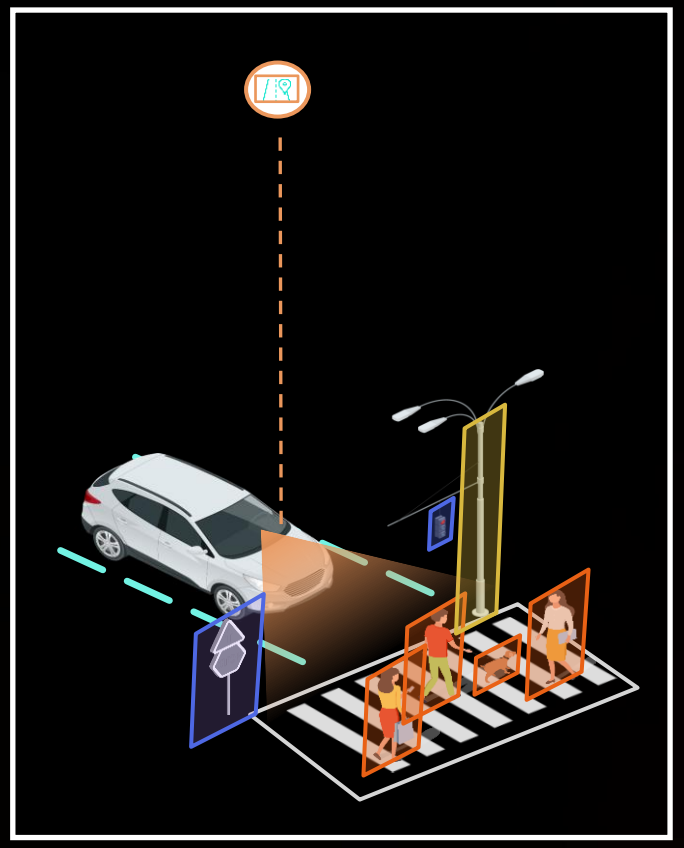
Harvesting

Aggregate – Align – Model



Roadbook





- Upload detected road elements (small compressed payload)
- Organic road traffic driven by human drivers



aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

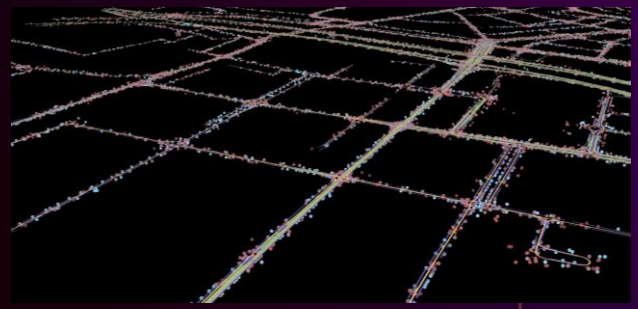
Aggregation



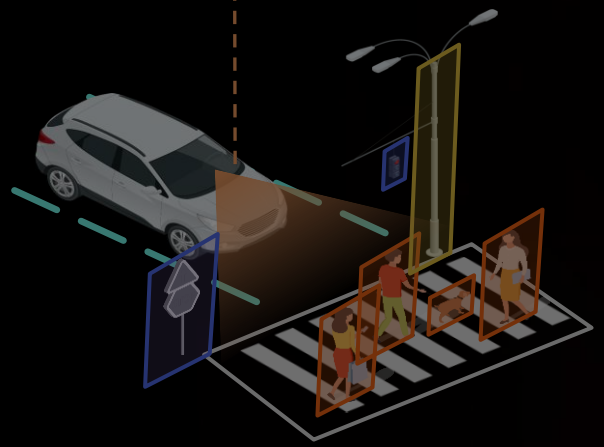
Aggregate – Align – Model



Roadbook







- Merge data from multiple drives per road
- Align the observations
- Infer accurate 3D road-structure geometry
- Infer road semantics & driving culture from human driver behavior
- Encode it as a detailed map – Roadbook

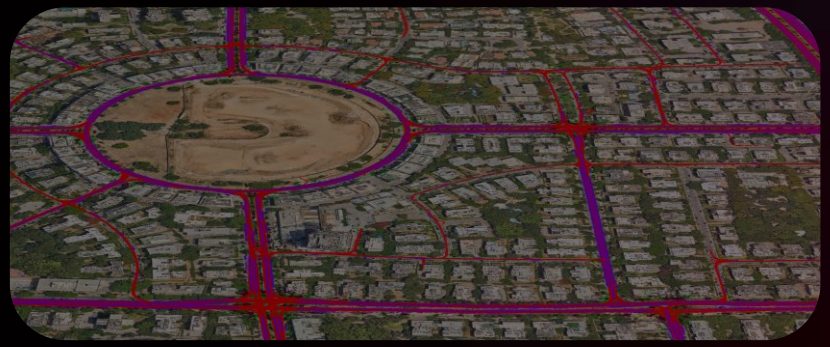


aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

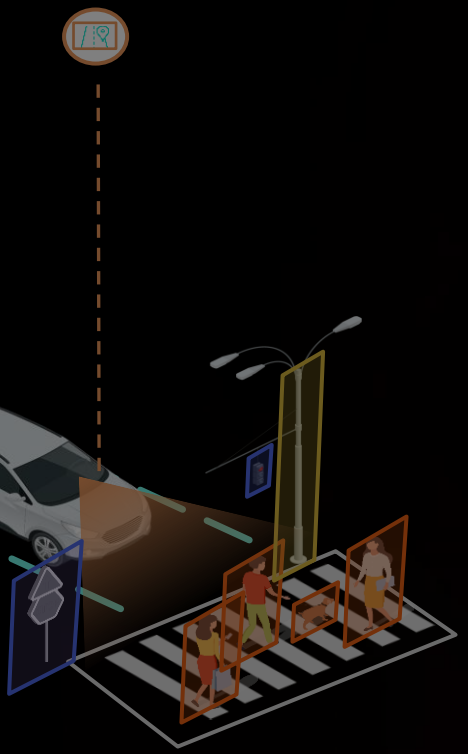
Localization

Aggregate – Align – Model

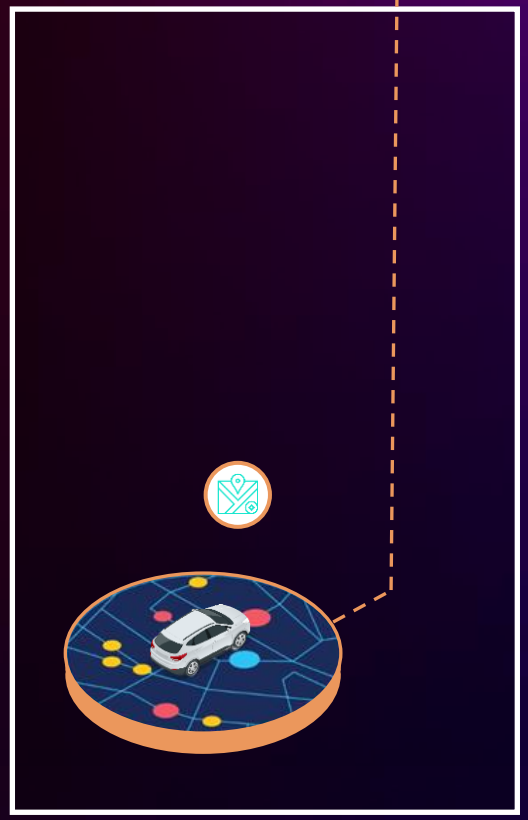


Roadbook



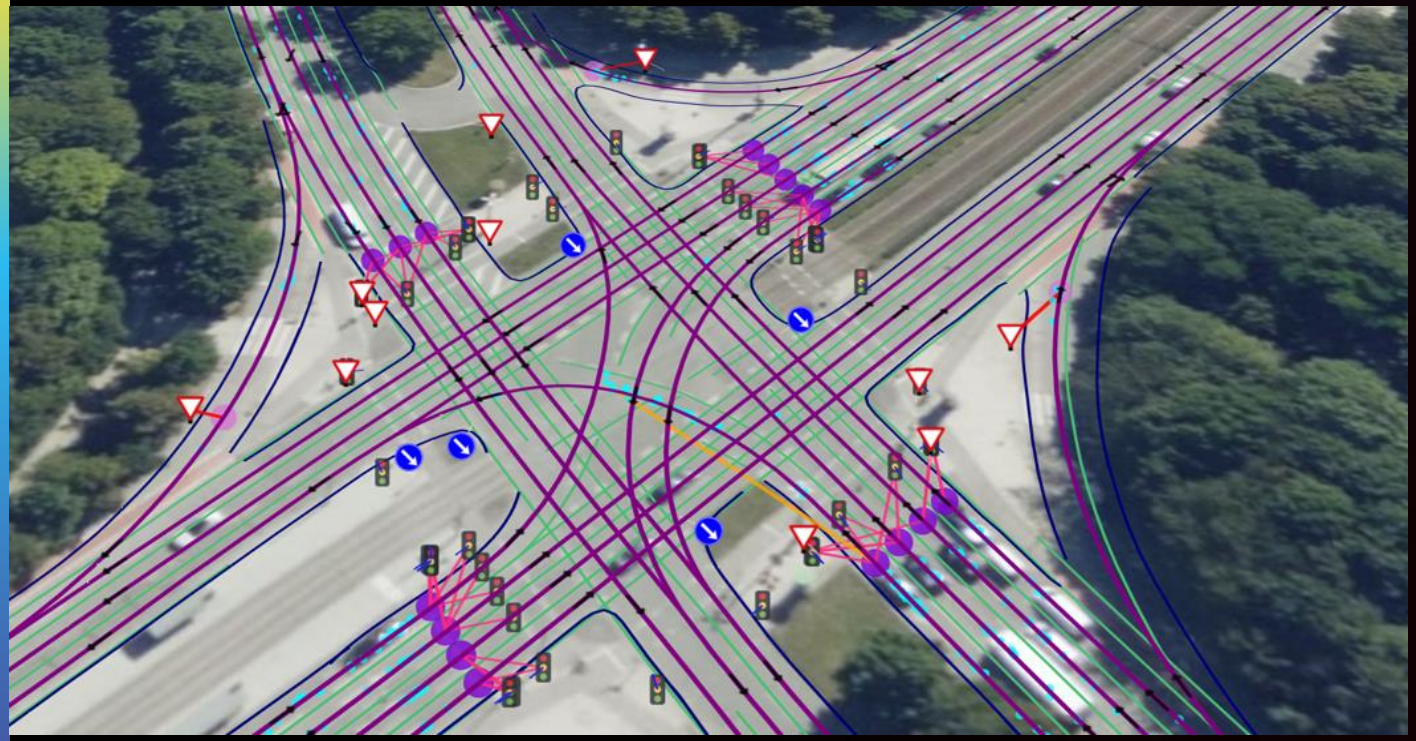


- Applications download tiles of the map on demand
- Localize into map – Position and pose to within cm-level accuracy



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Roadbook



mapbox

Road and lane structure

- Topology & geometry

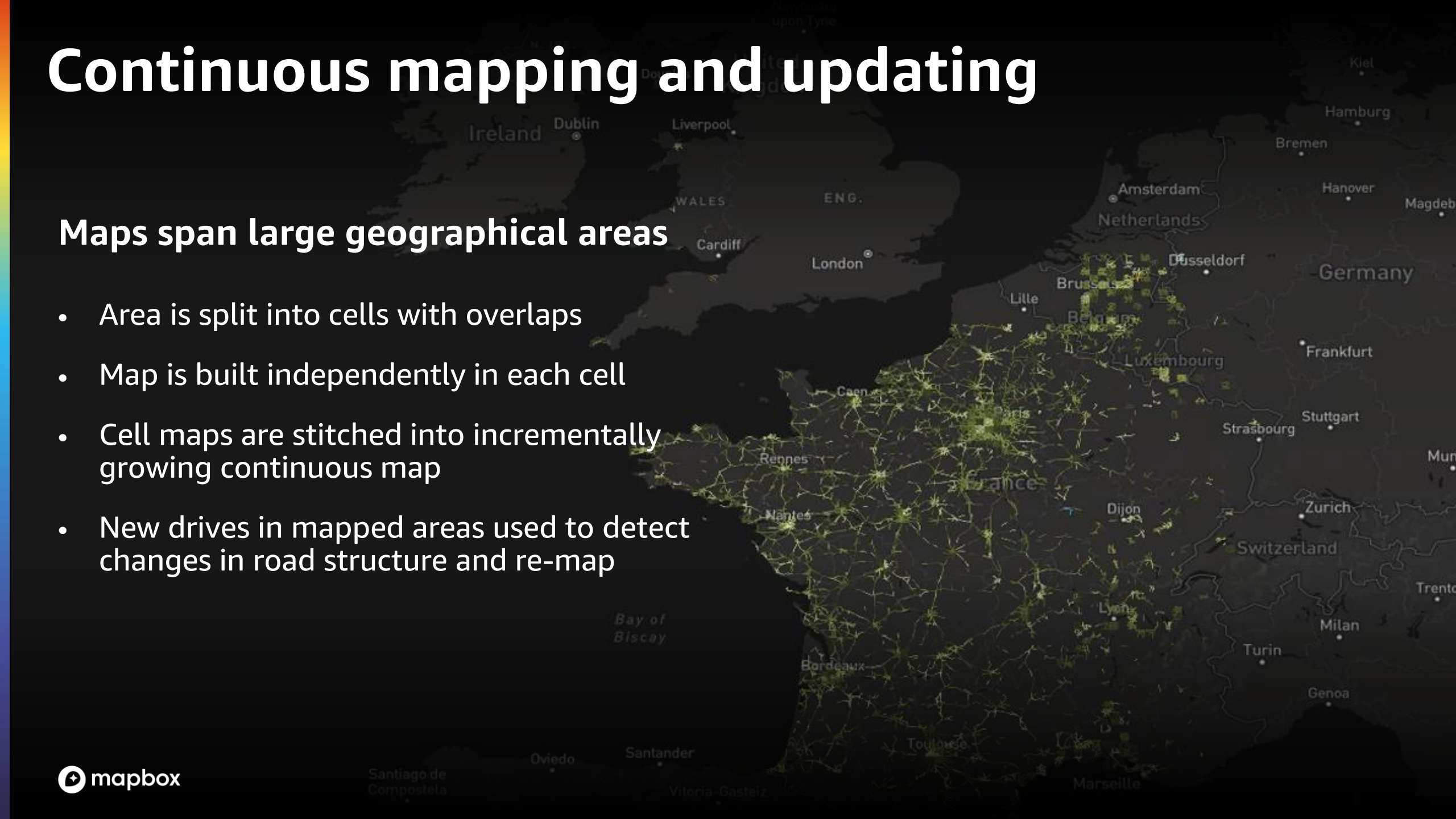
Road semantics

- Sign – Traffic light – Stop point – Lane
- Common and legal speed per lane
- Right of way information
- And more . . .

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Continuous mapping and updating



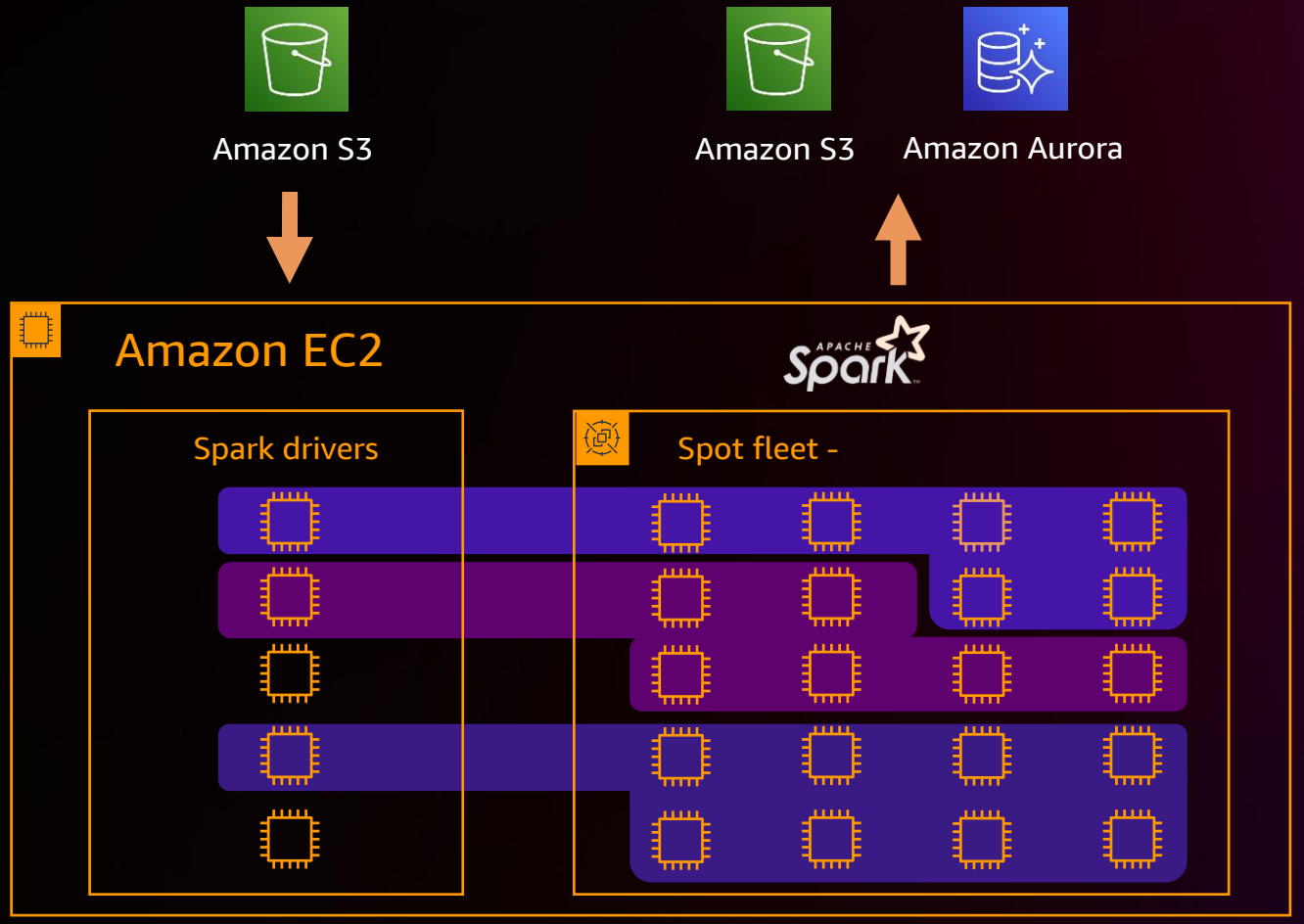
Maps span large geographical areas

- Area is split into cells with overlaps
- Map is built independently in each cell
- Cell maps are stitched into incrementally growing continuous map
- New drives in mapped areas used to detect changes in road structure and re-map

mapbox

REM tech stack

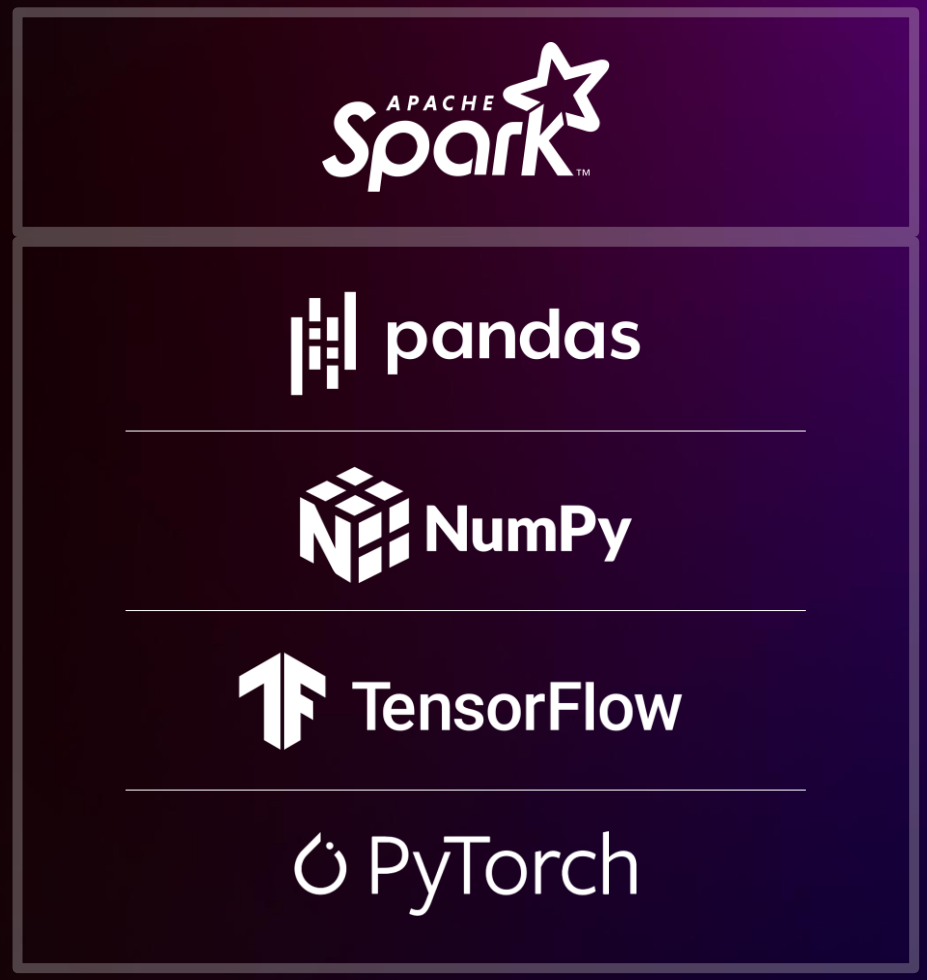
100K+ CPU cores and hundreds of map cells concurrently



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

REM tech stack – Map creation

Data manipulation →



Python functions & custom algorithms
Observation alignment
Clustering & modeling
Deep neural networks & AI

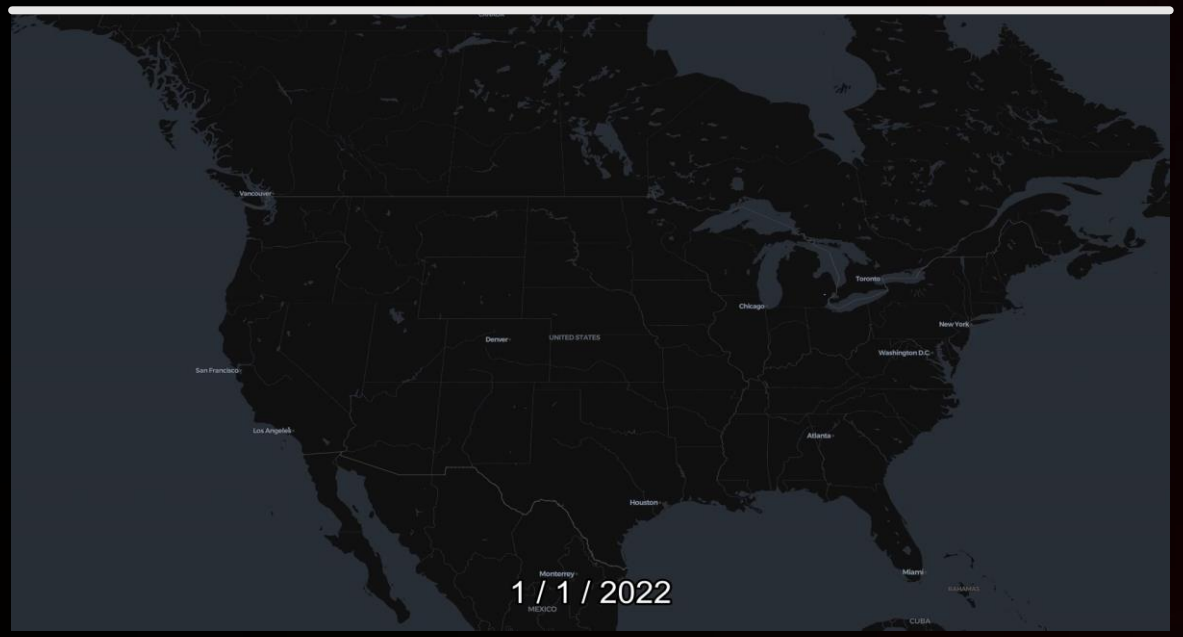
aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Scale at REM – Harvesting

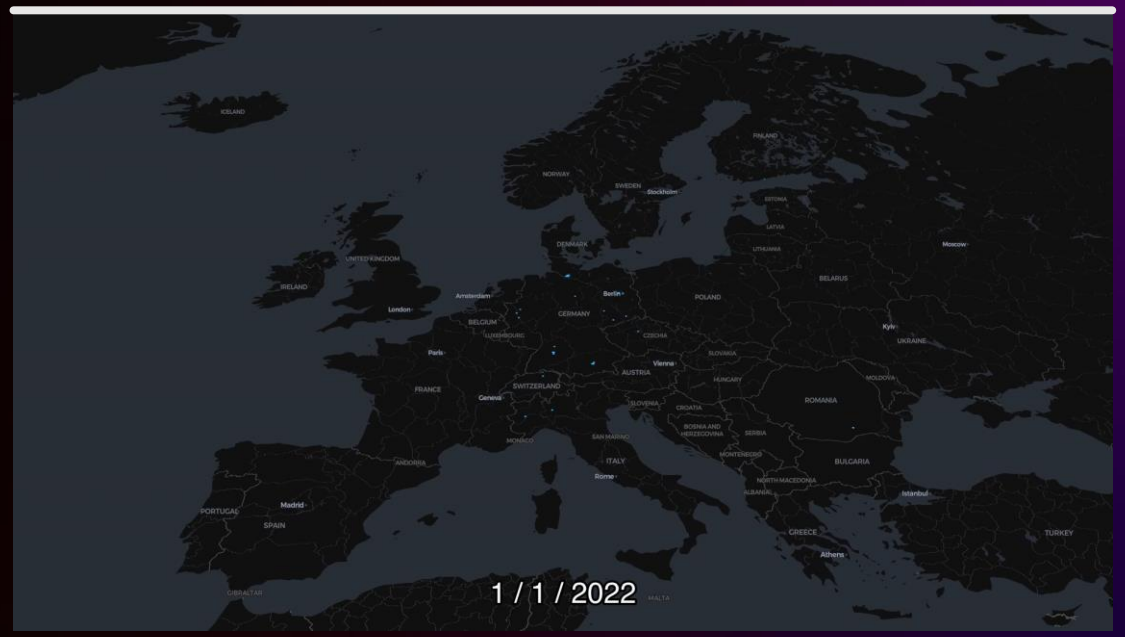
REM is a hyper growth activity

~14B KMs of road data collected as of July 2022

~69M KMs of road data analyzed per day



mapbox



mapbox

Pixel = 1,000+ drives

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Scale at REM – Map creation



Japan Highway road network (highways): ~15K km

North Germany inter-city road network: ~200K km

Europe inter-city road networks: ~2M km

- Europe + USA + China road networks: ~6M km
- Israel + multiple cities worldwide – all road network

Maps richer with features as AV and L2+ functions use more information for more complex functionality

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

REM R&D

Growing

- More features and improved map quality
- Testing at larger scale to ensure low MTBF

Fast paced

- AV development is rapid → demand for new features is frequent
- New map features pass from idea → implementation → production within weeks
- Algorithmic development is ongoing – We keep innovating to improve!

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Cost challenges

Running at REM scale is costly

Bulk of cost: **cpuh of EC2 instances**

KPI :

$$\text{average } \frac{\text{cpuh}}{\text{km}} \text{ of road lane}$$

Invest in cost-optimization activities before scale makes cost too high

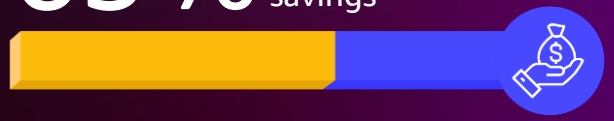
aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

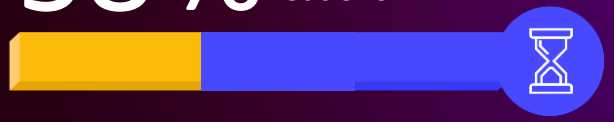
And then we met Granulate

Granulate automatically improves performance and reduces costs by optimizing OS and runtime resource management – with no R&D efforts

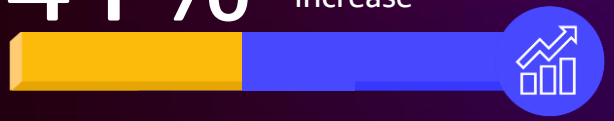
Up to **63%** compute cost savings



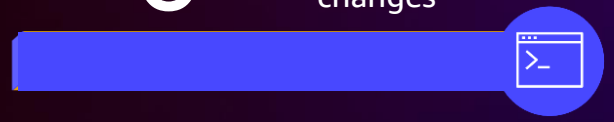
38% response time reduction



41% throughput increase



0 code changes



Autonomous, continuous workload optimization

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Granulate POC

1



PROFILING

24% improved performance w/in 3 weeks

2



LEARNING

Deployed to Spark machines (drivers & works) in AWS environment (3 weeks)

3



OPTIMIZING

Same map, same code, same correction – With Granulate, cpuh reduced by 45%

4



COST REDUCTION

45% cost reduction of EC2 on demand, EC2 Spot, and Amazon EBS volumes (each attached to EC2)

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Results

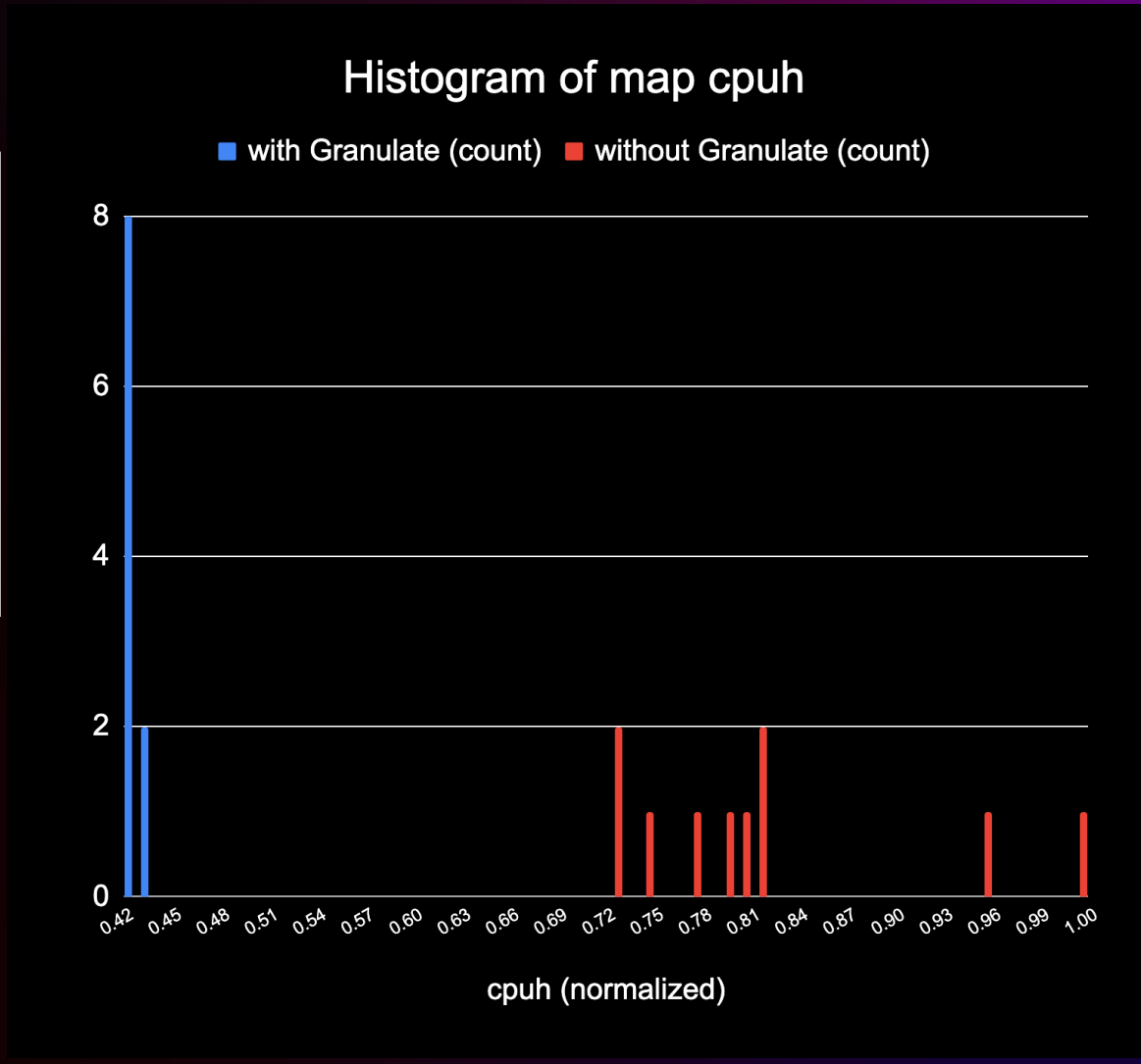
Ran:

10 repetitions without Granulate optimizing

10 repetitions with Granulate optimizing

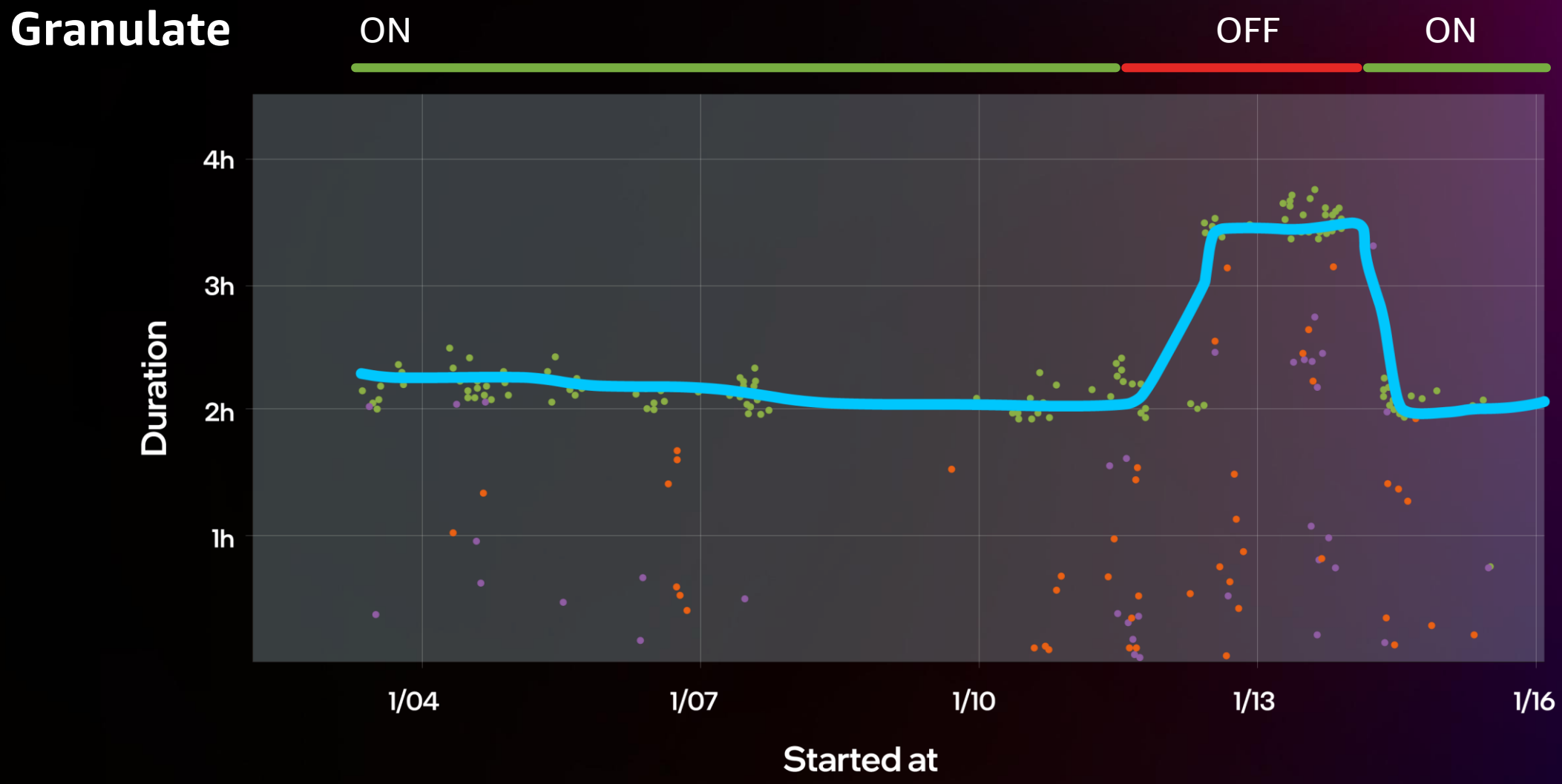
Same map. Same code. Same configuration (except Granulate)

~45% less cpuh used



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

An unexpected side result . . .



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Results – Beyond the cost savings

→ Supports fast innovation and pushing new features into production

→ Defer dedicated optimization efforts to a later stage

→ Developers focus on better maps for better driving experience and safety

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

habana® An Intel Company | intel® | GRANULATE™ An Intel Company

Intel-powered AWS instances and software optimizations empower developers and accelerate our customers' innovations

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!

Learn more at Intel booth #1617 or AI Neighborhood #2741



Please complete the session survey in the **mobile app**

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.