

Edge Computing for the Infrastructure

Victor Bahl

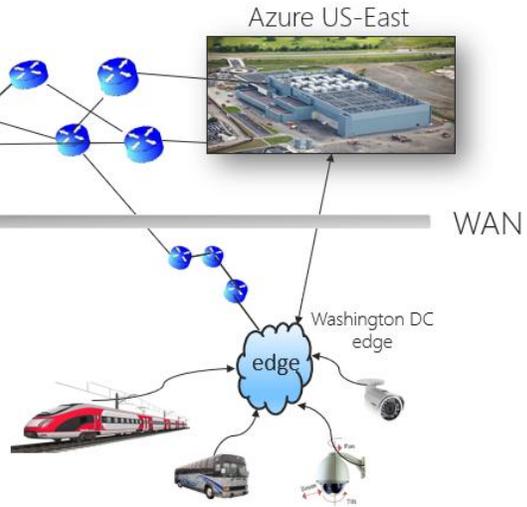
Technical Fellow - Research

Microsoft



@SuperBahl

3. where are we going?



Computer

COVER FEATURE EDGE COMPUTING

Real-Time Video Analytics: The Killer App for Edge Computing

Genesh Ananthanarayanan, Parasuram Bahl, Peter Bostik, Krishna Chintalapudi, Mathai Philipose, Lenin Ravindranath, and Sridhar Srinivas, Microsoft Research

Video analytics will drive a wide range of applications with great potential to impact society. A geographically distributed architecture of public clouds and edges that extend down to the cameras is the only feasible approach to meeting the strict real-time requirements of large-scale live video analytics.

According to a 2015 report by the Information Handling Division on the installed base for video surveillance equipment, there is a camera installed for every 20 people on the planet, with mature markets having a camera for every 8 people. The report predicts that the number of cameras will grow by 20 percent year over year for the next 5 years. Video analytics from these cameras are used for traffic control, surveillance, and security in both public and private venues, as well as consumer applications including digital assistants for real-time decisions.

We propose that a geographically distributed architecture of public clouds, private clusters, and edges that extend down to the cameras is the only approach that can meet the strict real-time requirements of large-scale video analytics which must address latency, bandwidth, and provisioning challenges.

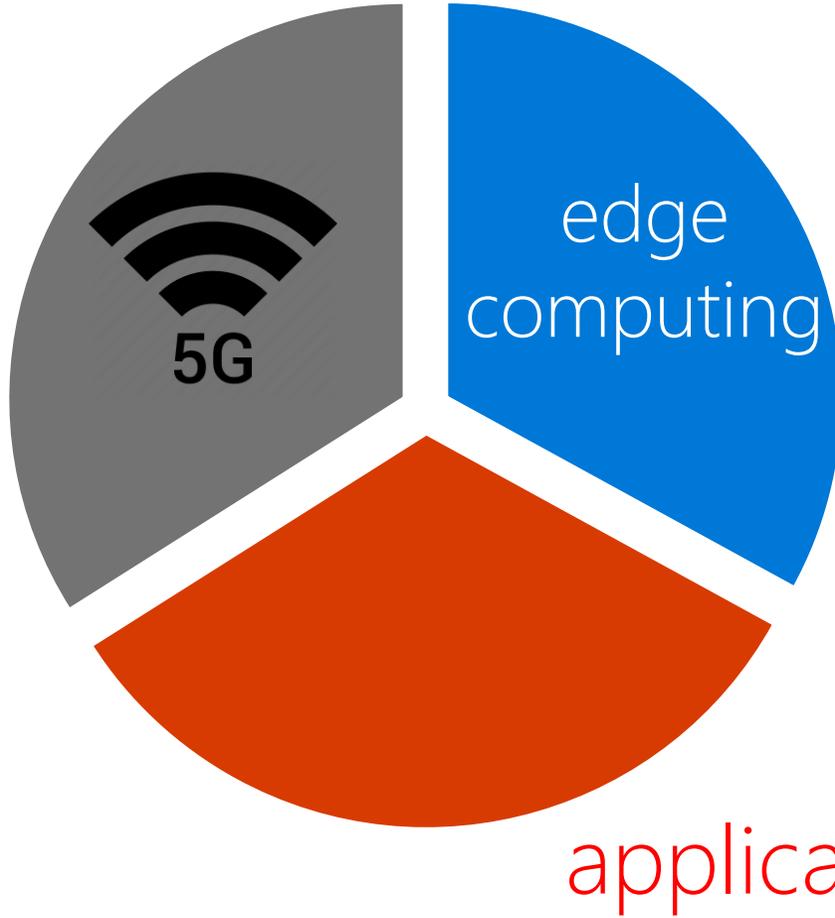
First, applications require very low latency when processing video because the output of the analytics is used to interact with humans (such as in augmented reality

accelerated) or to actuate some other system (such as traffic lights). Second, high-definition video requires large bandwidth—5 Mbps or even 25 Mbps for 4K video—and streaming a large number of video feeds directly to the cloud might not be feasible. When cameras are connected wirelessly, such as inside a car, the available uplink bandwidth is very limited. Finally, using compute capacity available on the camera itself allows for correspondingly lower provisioning (or usage) in the cloud. This also means that less interesting parts of the video can quickly be filtered out, for example, using motion-detection techniques, which dramatically reduces the bandwidth that needs to be provisioned.

Aside from low latency and efficient bandwidth usage, another major consideration for continuous video analytics is video processing: high compute cost. Because of high data volumes, compute demands, and latency requirements, cameras are the most challenging “blunt” in the Internet of Things. Thus, large-scale video analytics could well be edge computing’s “killer app.” Tapping into the

18 COMPUTER FALL 2016 OF THE IEEE COMPUTER SOCIETY

talk agenda



1. how did we get here?



2. what are we doing?



8K x 120 Hz x 10 bit streaming

> 440 Mbps file downloads

live video analytics & derivatives

augmented reality

Microsoft Azure

millions of servers

miles intra-DC fiber

2M

160+

data centers

54

Regions worldwide

100+

Tb data on backbone

- Available region
- ⊙ Announced region
- ◆ Availability Zones

$O(100K)$ network devices

$O(100M)$ lines of configuration

$O(1K)$ maintenance changes/day



Microsoft's data centers



Columbia river, hydro-electric power

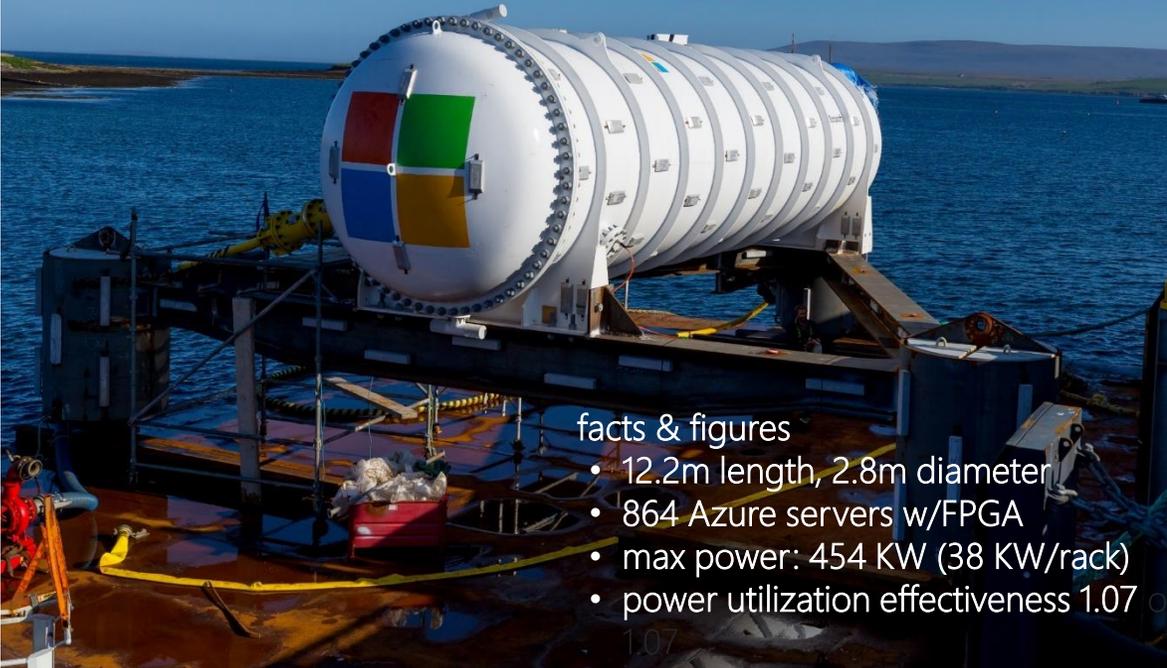


each facility is 8 MW in size, total of 64 MW



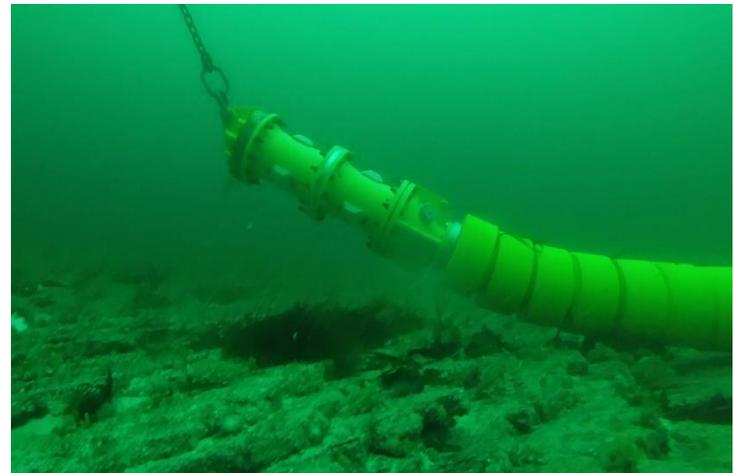
expanding rapidly, powered by wind farms

underwater data center



facts & figures

- 12.2m length, 2.8m diameter
- 864 Azure servers w/FPGA
- max power: 454 KW (38 KW/rack)
- power utilization effectiveness 1.07



Azure growing at a phenomenal rate

FY18: Intelligent Cloud: \$23 billion revenue (Azure grew 89% in Q4)



>94%
of Fortune 500 use
Microsoft Cloud

... in parallel we proposed edge computing

looking beyond cloud computing
October 29, 2008 in Bldg. 99, Redmond, WA

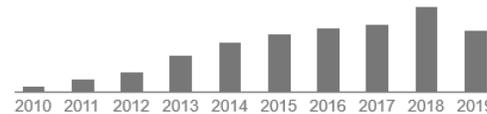
first paper



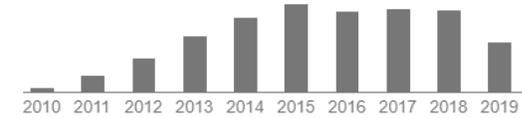
The Case for VM-Based Cloudlets in Mobile Computing

A new vision of mobile computing liberates mobile devices from severe resource constraints by enabling resource-intensive applications to leverage cloud computing free of WAN delays, jitter, congestion, and failures.

Cited by 3025



Cited by 2397



MAUI: Making Smartphones Last Longer with Code Offload

Eduardo Cuervo[†], Aruna Balasubramanian[‡], Dae-ki Cho^{*}, Alec Wolman[§], Stefan Saroiu[§], Ranveer Chandra[§], Paramvir Bah[§]
[†]Duke University, [‡]University of Massachusetts Amherst, ^{*}UCLA, [§]Microsoft Research

ABSTRACT

This paper presents MAUI, a system that enables fine-grained energy-aware offload of mobile code to the infrastructure. Previous approaches to these problems either relied heavily on programmer support to partition an application, or they were coarse-grained requiring full process (or full VM) migration. MAUI uses the benefits of a managed code environment to offer the best of both worlds:

Given the tremendous size of the mobile handset market, solving the energy impediment has quickly become the mobile industry's foremost challenge [14].

One popular technique to reduce the energy needs of mobile devices is *remote execution*: applications can take advantage of the resource-rich infrastructure by delegating code execution to remote servers. For the last two decades, there have been many attempts to make mobile devices use remote execution to improve perfor-

first article



Why a Cloudlet Beats the Cloud for Mobile Apps



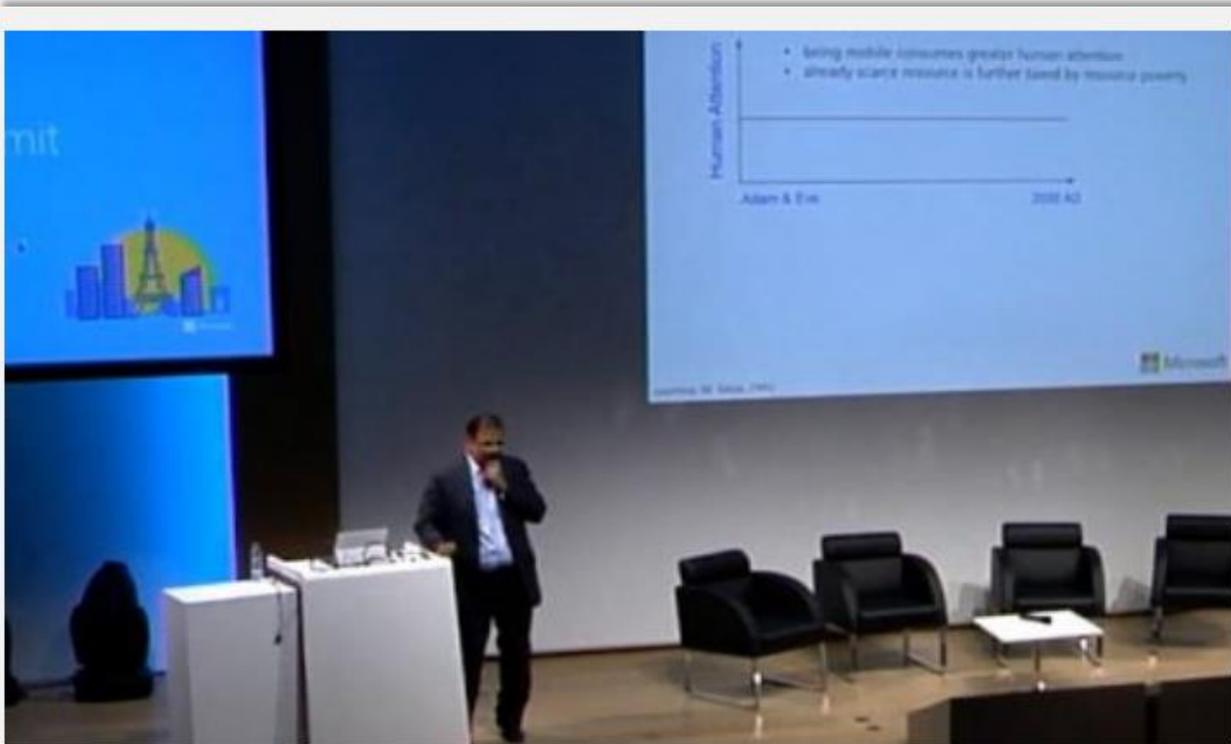
Posted on December 13, 2009 by lewisshepherd



first live demo



evangelization



Emergence of Micro-datacenters (Cloudlets/Edges for Mobile Computing)

May 13, 2015

Speakers: Victor Bahl

Affiliation: Microsoft Research

SEP 3, 2015 4:07 AM PDT

Microsoft researcher: Why Micro Datacenters really matter to mobile's future

And why enterprise IT pros should be getting ready for them



By Bob Brown

Network World | SEP 3, 2015 4:07 AM PDT

NETWORKWORLD
FROM IDG



Microsoft Research distinguished scientist Victor Bahl has been spreading the word about Micro Datacenters, also known by the adorable name "cloudlets," as a key concept for optimizing the performance and usefulness of mobile and other networked devices via the cloud. Service providers have embraced this vision most strongly from the start, but it won't be long before enterprise IT pros will likely do the same, Bahl says.

six PhD thesis in systems so far ...

- **Kevin Hsieh**, *Low-Latency, Low-Cost Machine Learning Systems on Large-Scale, Highly-Distributed Data*, Carnegie Mellon University (September 2019)
- **Chien-Chun ("Michael") Hung**, *Resource scheduling in Geo-distributed Computing*, University of Southern California (December 2017)
- **Shubham Jain**, *Design of Inertial & Camera Sensing for Smart Intersections*, Rutgers University, (August 2017)
- **Grassi, Giulio**, *Connected cars: A computing resource for smart cities*, Université Pierre-et-Marie-Curie, Paris (October 2017)
- **Yuan (Tiffany) Chen**, *Interactive Object Recognition and Search over Mobile Video*, Massachusetts Institute of Technology (June 2017)
- **Robert LiKamwa**, *Vision Sensing Pipeline for Efficiency & Privacy*, Rice University (July 2016)

more coming



best paper award



highlights

lots of media coverage



Human and computer vision unite to help Microsoft engineers stop traffic deaths

BY LISA STIFFLER on June 1, 2017 at 6:00 am



Intelligent cameras can put an end to always-on surveillance

Many cities are packed with cameras pointlessly recording everything they see, but smart algorithms could allow them to keep only footage that matters



Microsoft looks to stop bike crashes before they happen, testing Minority Report-style predictive intelligence

and many more

publications



along the way IoT hit
creating a big opportunity ...

Microsoft Pledges \$5 Billion IoT, Edge Push



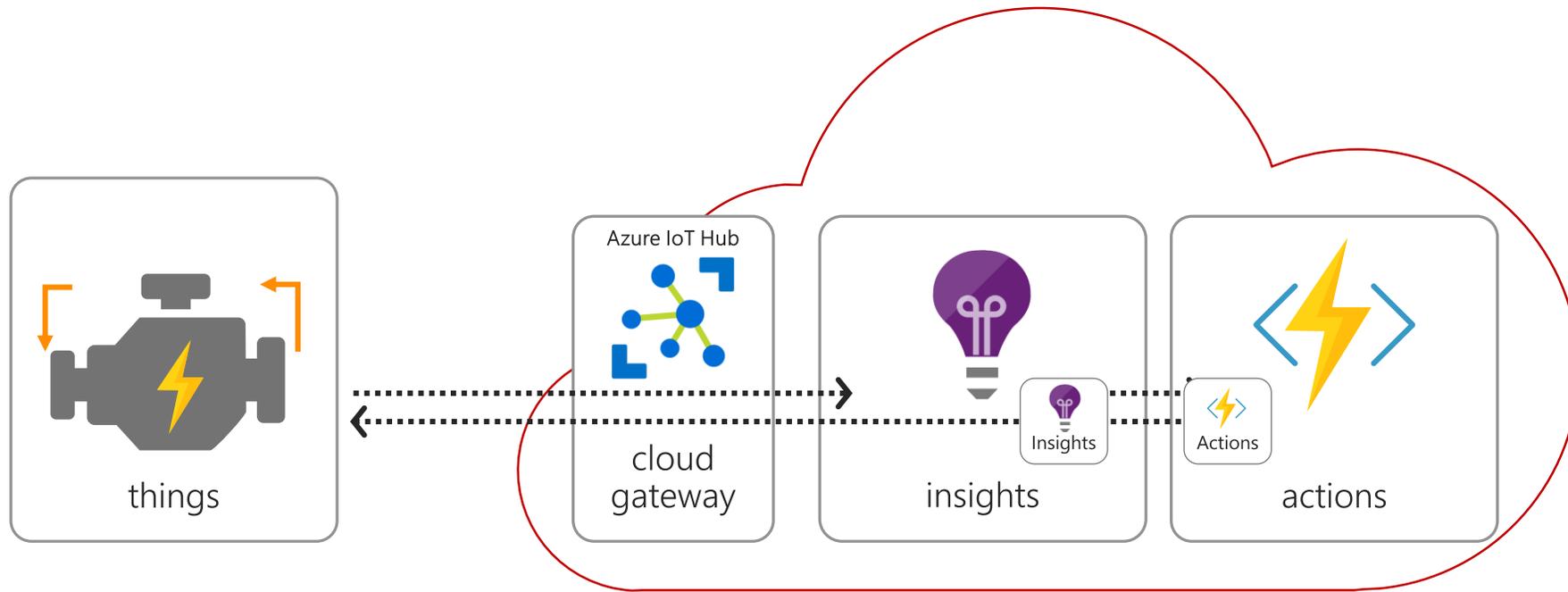
Jessica Lyons Hardcastle | Managing Editor

April 5, 2018 10:38 AM

Share this article:



IoT app pattern & edge



cloud services
at the edge

Azure ML, Azure Stream
Analytics, Azure Functions,
custom

manage from
the cloud

devices & services from
Azure Portal

flexible
connectivity

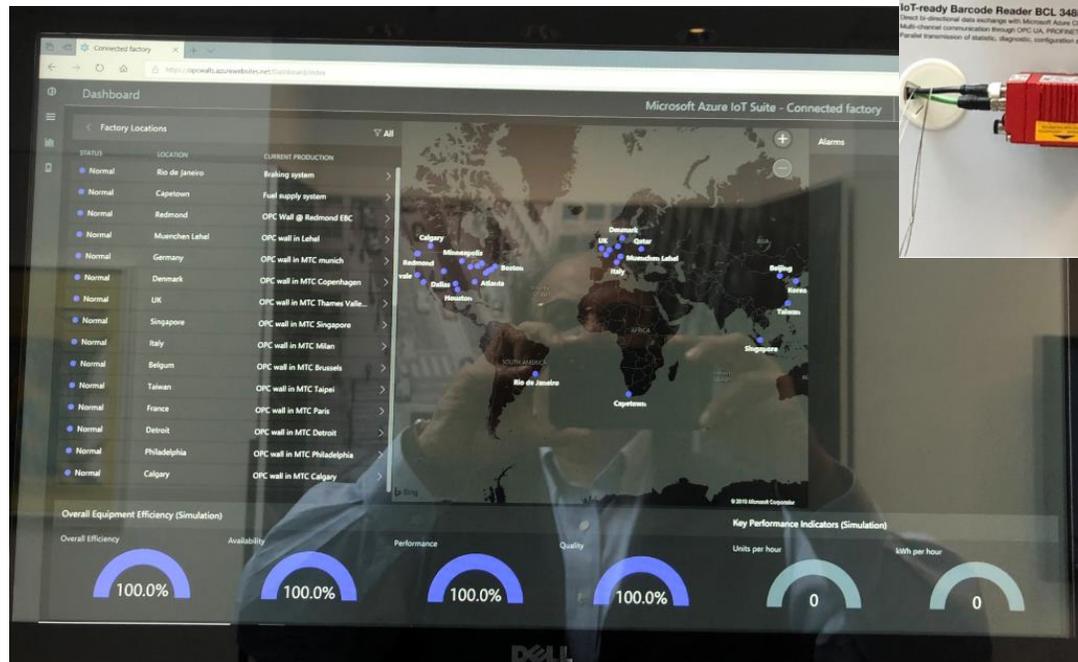
intermittent, low, or no
connectivity

reduced latency & cost

bring compute to the data,
reduced bandwidth cost

Azure IoT Edge

devices connect to cloud via Azure IoT Edge managed from the Azure portal; configure system, monitor health, & install updates from the cloud



Azure Stack Edge

announced Sept. 24, 2018 & Nov. 4, 2019

managed AI appliance



Commercial series

Tailored for most commercial scenarios, such as retail stores and datacenters

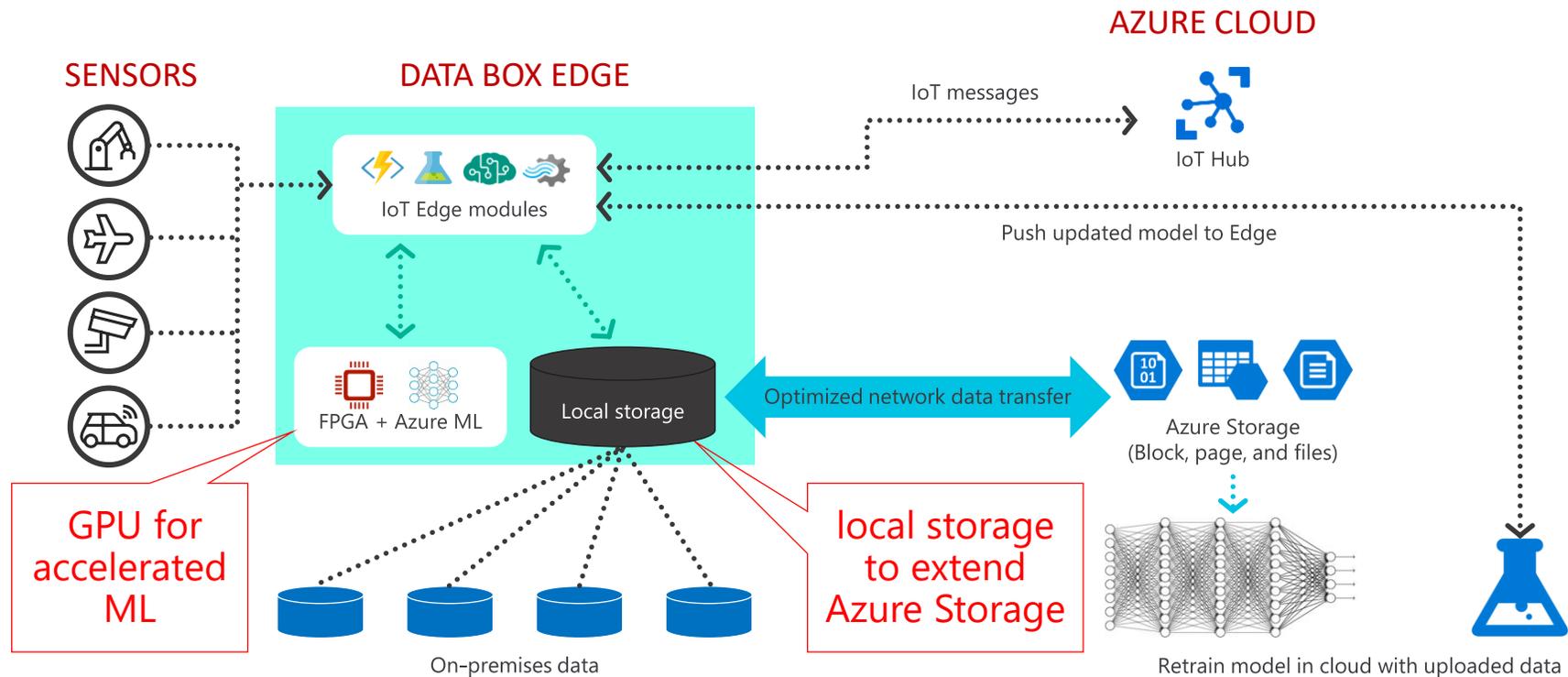


NEW Rugged series

Tailored for harsh environmental or field conditions, such as in defense, disaster relief, geological surveys, and energy

clients include Hong Kong Stock Exchange, KPMG Norway, Airbus Defense & Space, ...

Azure Stack Edge with Azure IoT Edge



Shell: safety at a gas station



Alaska DoT safe non-iced roads



Exxon: preventive maintenance



BMW automated factory



aha moment! (2014-15)

THE WALL STREET JOURNAL.
China's 100 Million Surveillance Cameras



a  for every 8 people in the US & for every 29 people worldwide!

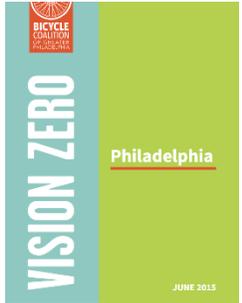
theguardian

You're being watched: there's one CCTV camera for every 32 people in UK

with cloud computing, it's the golden era for ML, computer vision, & AI

potential to impact science, society & business

cities all over North America are embracing it



Video Analytics Towards Vision Zero



BY FRANZ LOEWENHERZ, VICTOR BAHL, PH.D., AND YINHAI WANG, PH.D.

For young people below the age of 35, motor vehicle crashes are the leading cause of death in the United States. In 2015, collisions resulted in 35,092 deaths and 2.4 million injuries. More than 1,100 children under the age of 15 were killed. The 7.2 percent increase in traffic fatalities from 2014 to 2015 represents the greatest percentage increase in nearly 50 years.¹ Yet despite the massive death toll, work to prevent traffic fatalities has been woefully lacking.

Many governmental agencies continue to rely on traditional traffic safety approaches. They intervene only after enough police crash reports are filed to trigger a High Crash Corridor designation. This reactive approach to preventing crash recurrence has well-documented limitations:

- At most locations, the number of crashes is very small and subject to chance variations;
- Not all crashes are reported and the level of reporting is uneven regarding the type of road users involved, the exact location, and the severity of injuries;
- Numerous "close calls" go undocumented; and
- Many years of crash data are typically required to develop an understanding of the situation.²

Given these trends, and the crash analysis tools presently employed, how will jurisdictions achieve what all of us want: zero fatalities and serious injuries on our roadways? That's the goal of Vision Zero,

what is one of the leading causes of death worldwide.³ It calls on government agencies to be proactive, identify risks, and take steps to prevent injuries on our roadways. Vision Zero encourages us to imagine a future in which we do not need to wait for crashes to occur in order to prevent others from happening.⁴

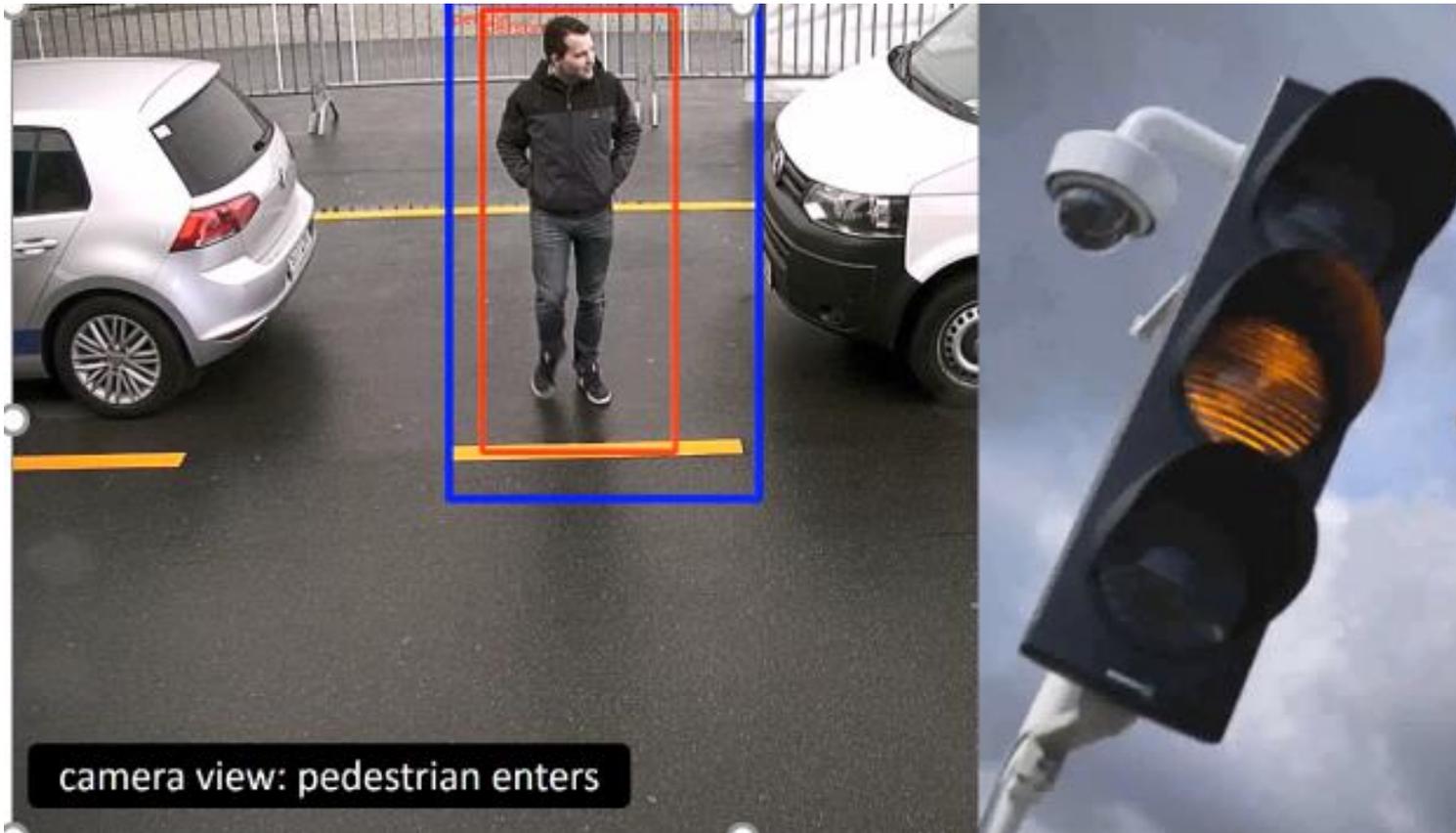
Solutions for a Safer World

Although traffic collisions can happen anywhere, there are often early warning signals in the form of conflicts or near-collision events at specific locations. These are recurring instances where a car abruptly stops because a bicycle veered in front of it, a pedestrian steps into the path of a bicyclist, or one bicyclist or car passes by another or a static object at very close spacing. These surrogate warning indicators – observable non-crash traffic conflict events – provide insight into when, where, and why crashes are most likely to occur. Understanding the root causes for near-collision events could enable local governments to take proactive, shes.⁵

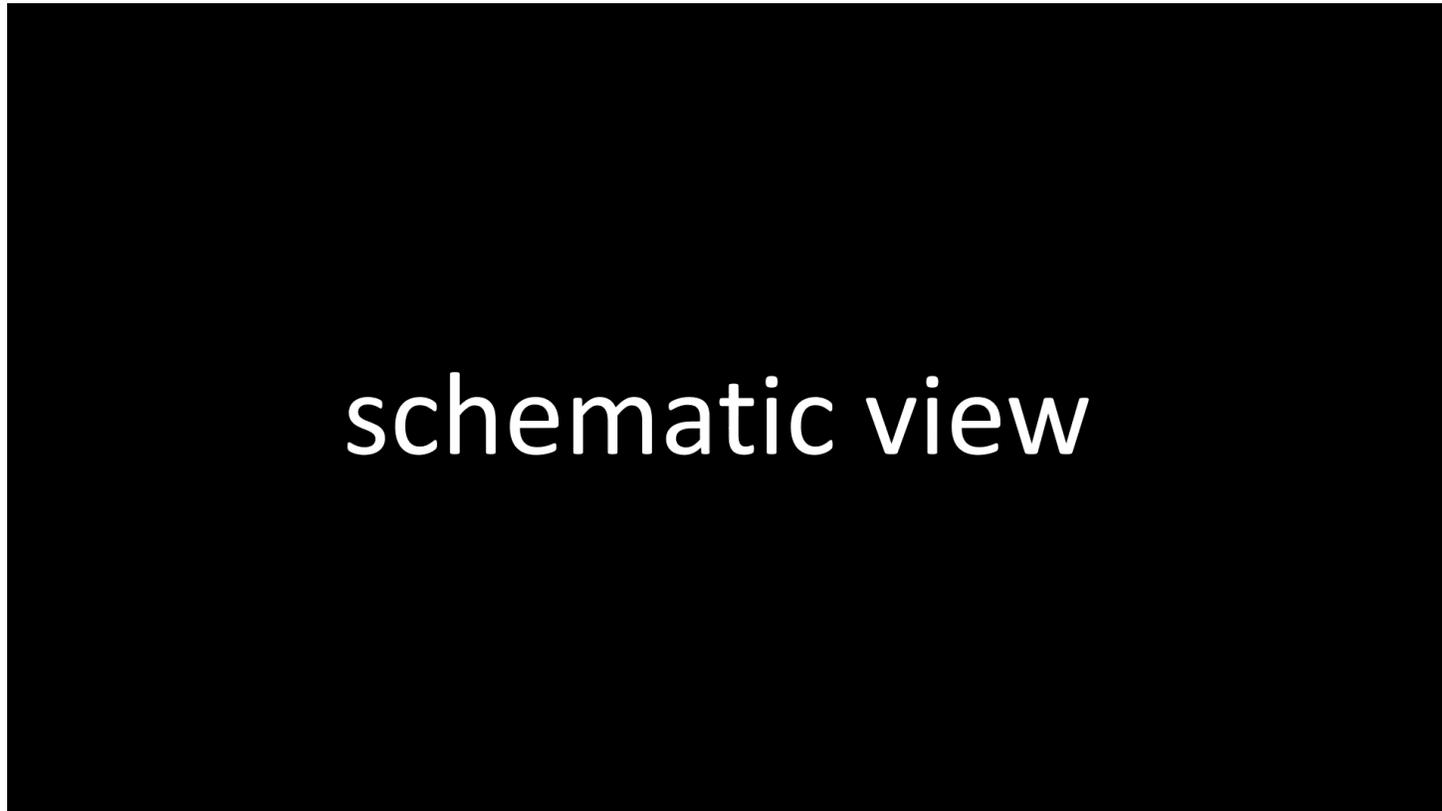
Microsoft + Bellevue Vision Zero Collaboration 2017 - 2019



systems demo-ed in 2016



live demos in Hannover Messe 2016



edge AI everywhere ... for good



AI for social good (Feb. 2019)



why focus was on intersections?

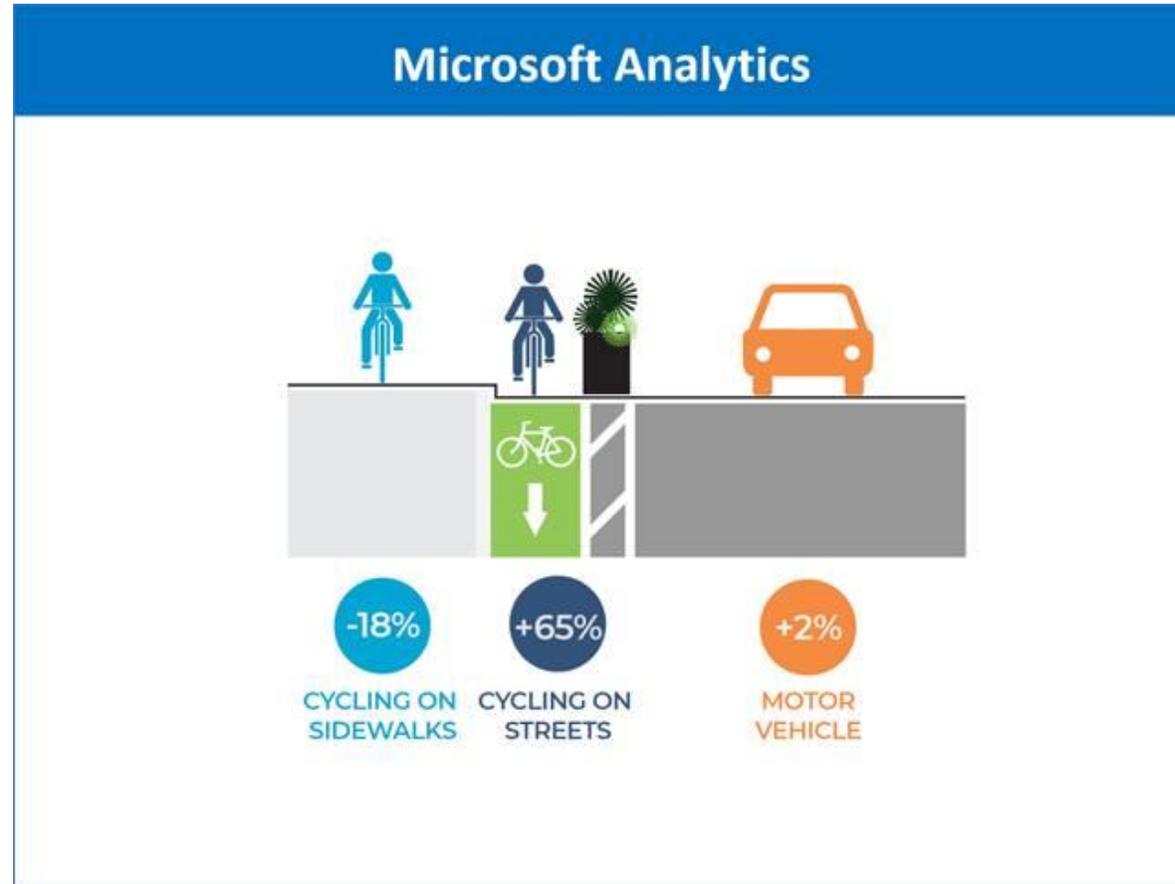
~20% fatal crashes happen at intersections
~50% of all crashes happen at intersections

and if this is not enough to motivate you, consider the business potential ...

United States has ~330,000 intersections with traffic sig

RANK	LEADING CAUSE	%
1	Ischaemic heart disease	12.2
2	Cerebrovascular disease	9.7
3	Lower respiratory infections	7.0
4	Chronic obstructive pulmonary disease	5.1
5	Diarrhoeal diseases	3.6
6	HIV/AIDS	3.5
7	Tuberculosis	2.5
8	Trachea, bronchus, lung cancers	2.3
9	Road traffic injuries	2.2
10	Prematurity and low birth weight	2.0
11	Neonatal infections and other	1.9
12	Diabetes mellitus	1.9
13	Malaria	1.7
14	Hypertensive heart disease	1.7
15	Birth asphyxia and birth trauma	1.5
16	Self-inflicted injuries	1.4
17	Stomach cancer	1.4
18	Cirrhosis of the liver	1.3
19	Nephritis and nephrosis	1.3
20	Colon and rectum cancers	1.1

108th Avenue NE, Bellevue, Washington bike lane project



cycling on sidewalks has reduced, on-street bicycle usage has increased due to the bike lane, & the volume of motor vehicle traffic is unaffected.

global partners



press release by DC DOT



FOR IMMEDIATE RELEASE:
August 30, 2017

MEDIA CONTACTS:
LaToya Foster (EOM) – (202) 727-5011; latoya.foster@dc.gov
Terry Owens (DDOT) – (202) 763-8635; terry.owens@dc.gov

Mayor Bowser to Mobilize Residents in Vision Zero Crowdsourcing Partnership

Analysis of Traffic Camera Footage to Support Vision Zero

(WASHINGTON, DC) – Today, as part of Washington, DC’s Vision Zero initiative, Mayor Bowser announced the Video Analytics Towards Vision Zero project, a cutting-edge partnership between the District and Microsoft that will use video footage and crowdsourcing to prevent traffic accidents. The project will tap artificial intelligence and new technologies to analyze traffic camera video footage and use near-miss collisions to predict where crashes are likely to occur in the future.

“Using video analytics to achieve Vision Zero is one more way we are building a smarter, safer, stronger DC,” said Mayor Bowser. *“Residents know traffic issues in their neighborhoods better than anyone, and now we will be able to leverage their knowledge with our existing camera infrastructure in order to prevent crashes and injuries before they occur.”*

The District has over 130 closed circuit television cameras around the city that are used to observe traffic conditions at intersections, and as part of this project, Microsoft has developed a crowdsourcing platform that will allow members of the public to review video footage and use tracking tools to identify movements and objects. People will be able to identify objects such as pedestrians, bicyclists, drivers, and motor-vehicles. The feedback will be used as part of a process where people will teach computers how to tell the difference between different movements and modes of transportation. Ultimately, instead of a person watching hours of video, computer algorithms will be able to analyze millions of hours of footage.

“The video analytics project will help us identify potential hazards at intersections throughout the District. Traditionally, we have used crash data for this purpose, but this new approach will help DDOT detect problems before crashes happen and before anyone is injured,” explained DDOT Interim Director Jeff Marootian.

“Computer vision algorithms applied to video feeds from traffic cameras have a huge potential of improving traffic flow and reducing traffic crashes and fatalities. We are working diligently on this because we truly believe the societal impact will be significant,” said Microsoft Research Distinguished Scientist Victor Bahl.

Residents can participate in the crowdsourcing platform by visiting the program website: <http://www.ite.org/visionzero/videoanalytics/>.

Vision Zero is the District’s plan to eliminate fatalities and serious injuries to people walking, biking, and driving within Washington, DC by the year 2024. To learn more about Washington, DC’s Vision Zero initiative, visit ddot.dc.gov/page/vision-zero-initiative.

###

Social Media:
Mayor Bowser Twitter: [@MayorBowser](https://twitter.com/MayorBowser)
Mayor Bowser Instagram: [@Mayor_Bowser](https://www.instagram.com/Mayor_Bowser)
Mayor Bowser Facebook: facebook.com/MayorMurielBowser
Mayor Bowser Website: mayor.dc.gov

mayor's challenge award to Bellevue

Safer Cities, Safer People Award



U.S. Department of Transportation
MAYORS' CHALLENGE



Winner!
Bellevue, Washington

Bellevue, WA, pursued a range of data collection activities during the Mayors' Challenge to identify barriers to bicycling and walking, prioritize improvements, and guide investments. In February 2015, the Bellevue City Council introduced the Pedestrian and Bicycle Implementation Initiative (PBII) to improve safety for people of all ages and abilities who walk and bike in Bellevue. Using data collected from online sources, key-pad polling at public meetings, automated bicycle and pedestrian counters, and traffic camera videos, the PBII team identified barriers to walking and bicycling and developed a \$4.8M Bicycle Rapid Implementation Program (BRIP) budget proposal to guide citywide investments through 2019. The BRIP aims to expand the city's bicycle network from 42 miles to more than 70 miles of conventional bike lanes, separated lanes or off-street paths, and to complete four continuous, cross-city bicycle corridors.

Demonstrated Successes
Innovative Data Collection Techniques: Gather Real-Time and Long-Term Data with Public Input
 Throughout the PBII process, Bellevue has

emphasized understanding long-term trends and gathering feedback from people who walk and bike. Bellevue's PBII team:

- Conducted a longitudinal assessment from 2006–2015 of non-motorized collisions using the USDOT's Pedestrian and Bicycle Crash Analysis Tool (PBCAT) system;
- Gathered input using key-pad polling and comment cards at 20 public meetings and an open house that attracted 140 attendees; and
- Used online surveys to solicit public input at two stages in the BRIP development process;
 - Over 700 people placed more than 1,600 points in the first online map to identify locations that they felt were unsafe for walking and bicycling;
 - Over 120 people submitted more than 400 comments on conceptual designs for 52 proposed projects to make the pedestrian and bicycle systems safer.



Pedestrian and Bicycle project manager Franz Loewenherz (foreground) and Councilmember Lynne Robinson (center) lead a policy ride with local bicycle advocates in Downtown Bellevue.

MAYORS' CHALLENGE: CHALLENGE ACTIVITY 3 (GATHER DATA) 1



Video Analytics analyzes traffic camera video footage and uses near-miss collisions to predict where future crashes are likely to occur. Traffic engineers could then take corrective action to prevent them. File photo

Bellevue video analytics project receives safety award

Fri Aug 4th, 2017 3:44pm · BUSINESS



Achievements Award



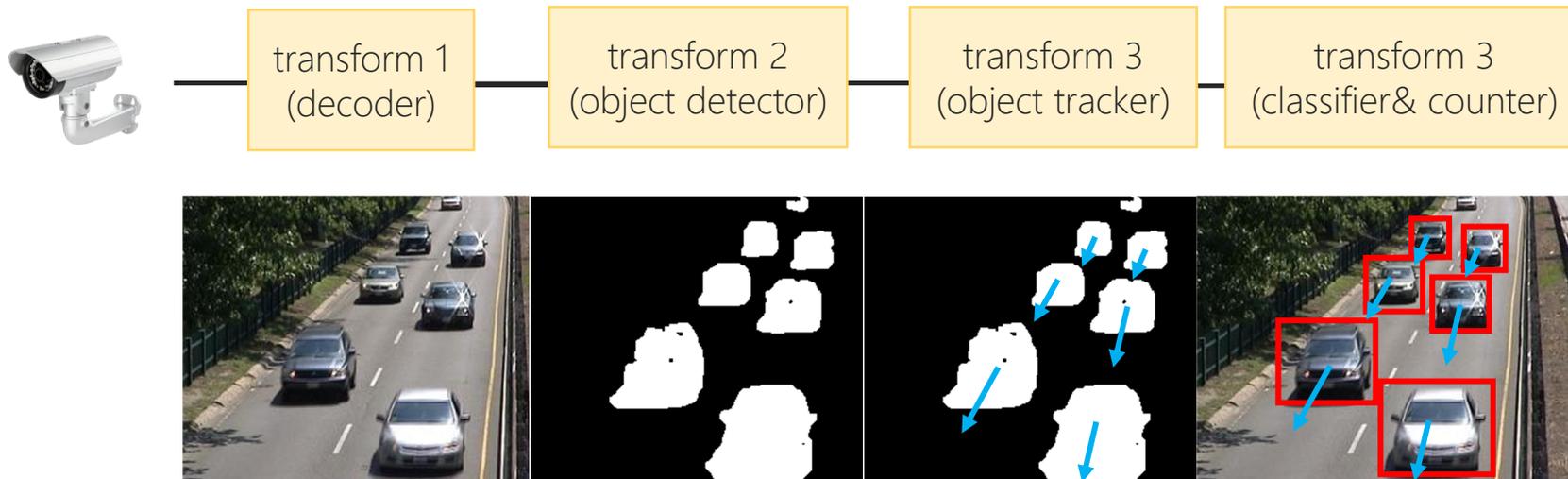
**U.S. Department
of Transportation**

a sampling of deeper dive...

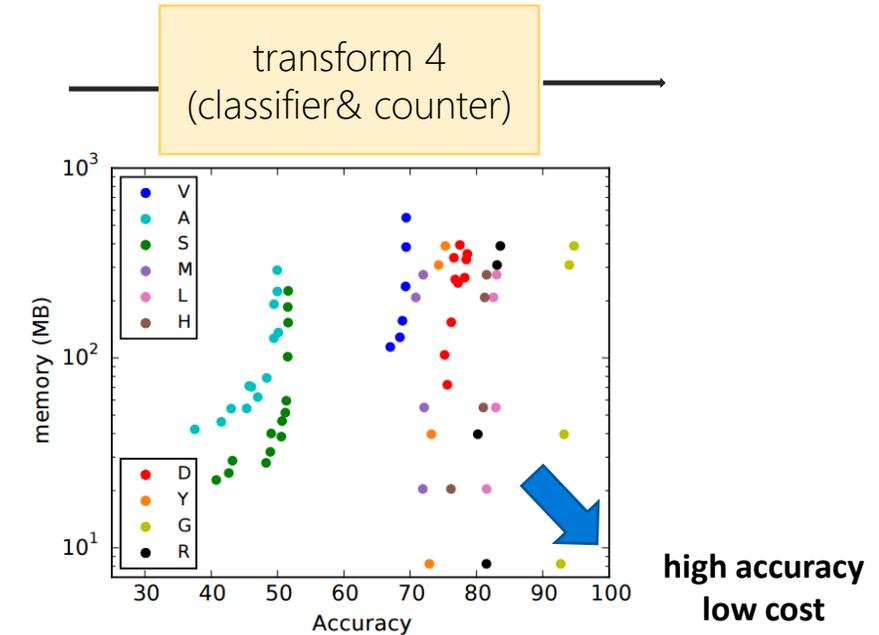
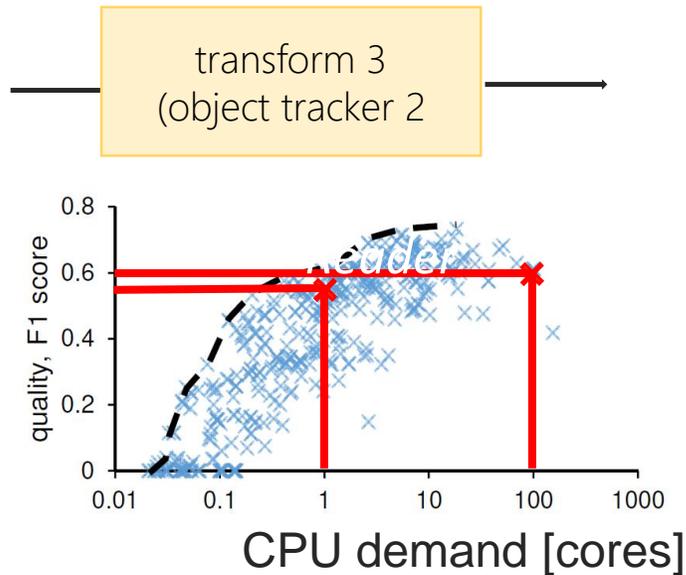
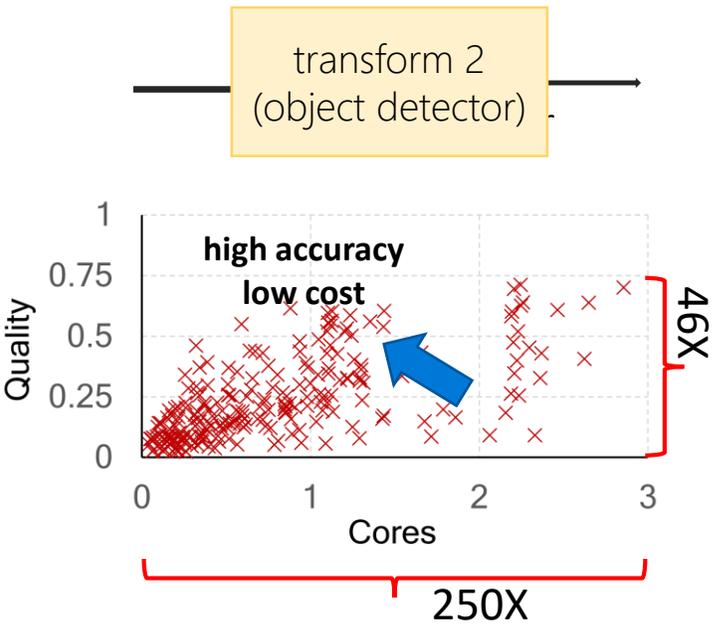
video query: pipeline of transforms

vision algorithms ("*transforms*") chained together
transforms implement specified interfaces

example: count the number of moving cars on a road segment



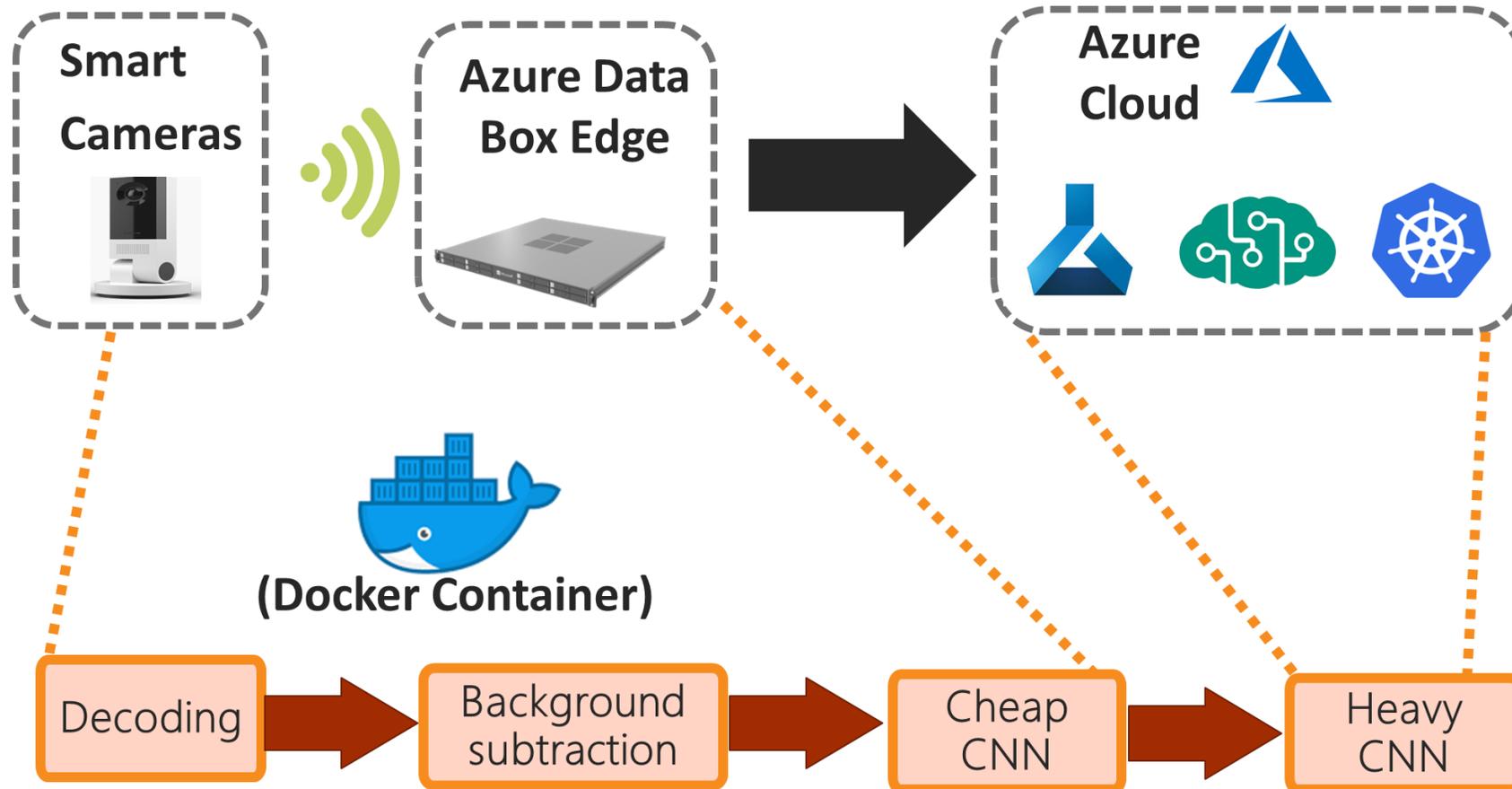
each stage has a resource/quality trade off



best car tracker^[1] — 1 fps on an 8-core CPU
DNN for object classification^[2] — 30GFlops

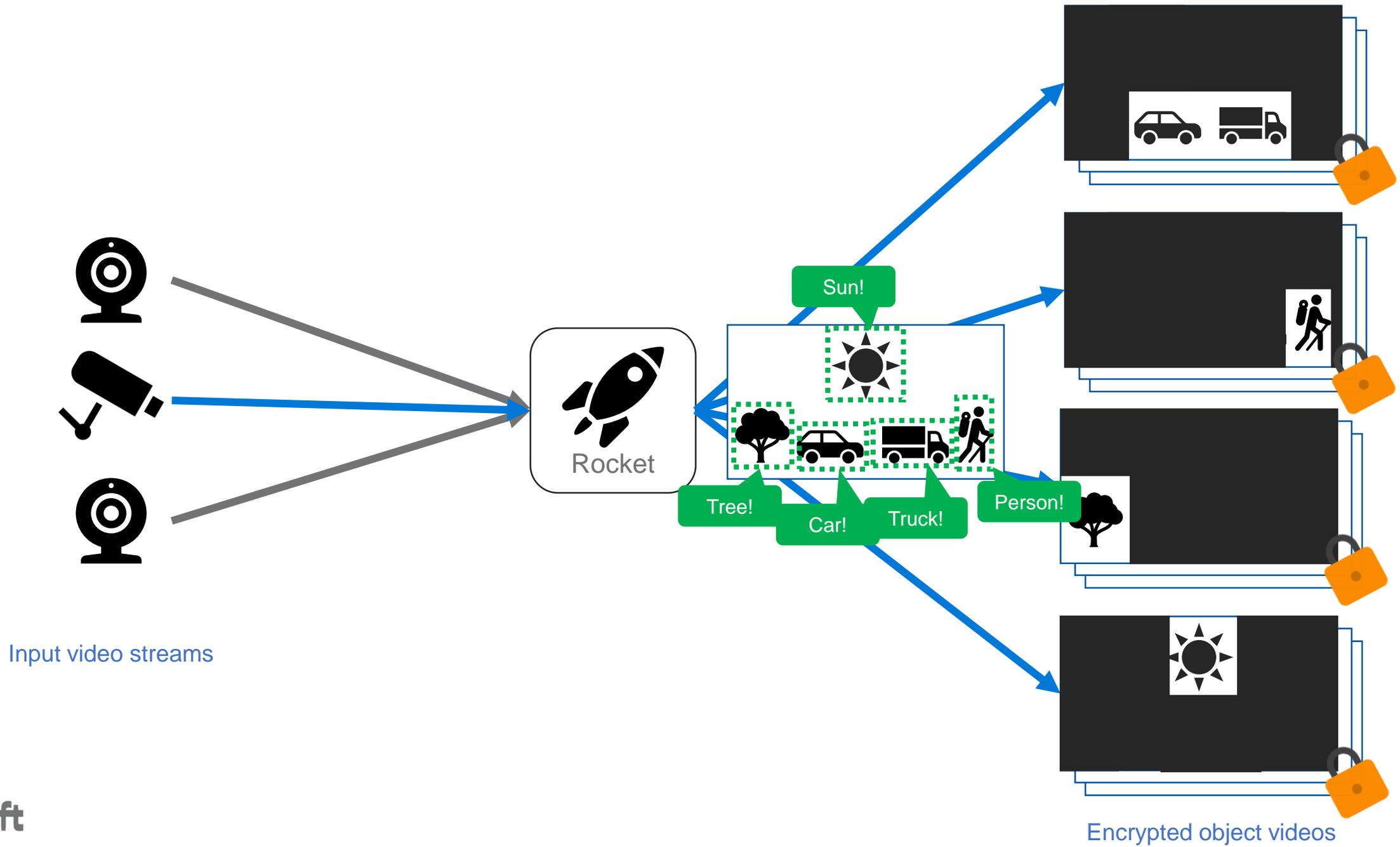
no one plan is uniformly the best...
best plan is dependent on the camera, lighting, track direction, object color, ...

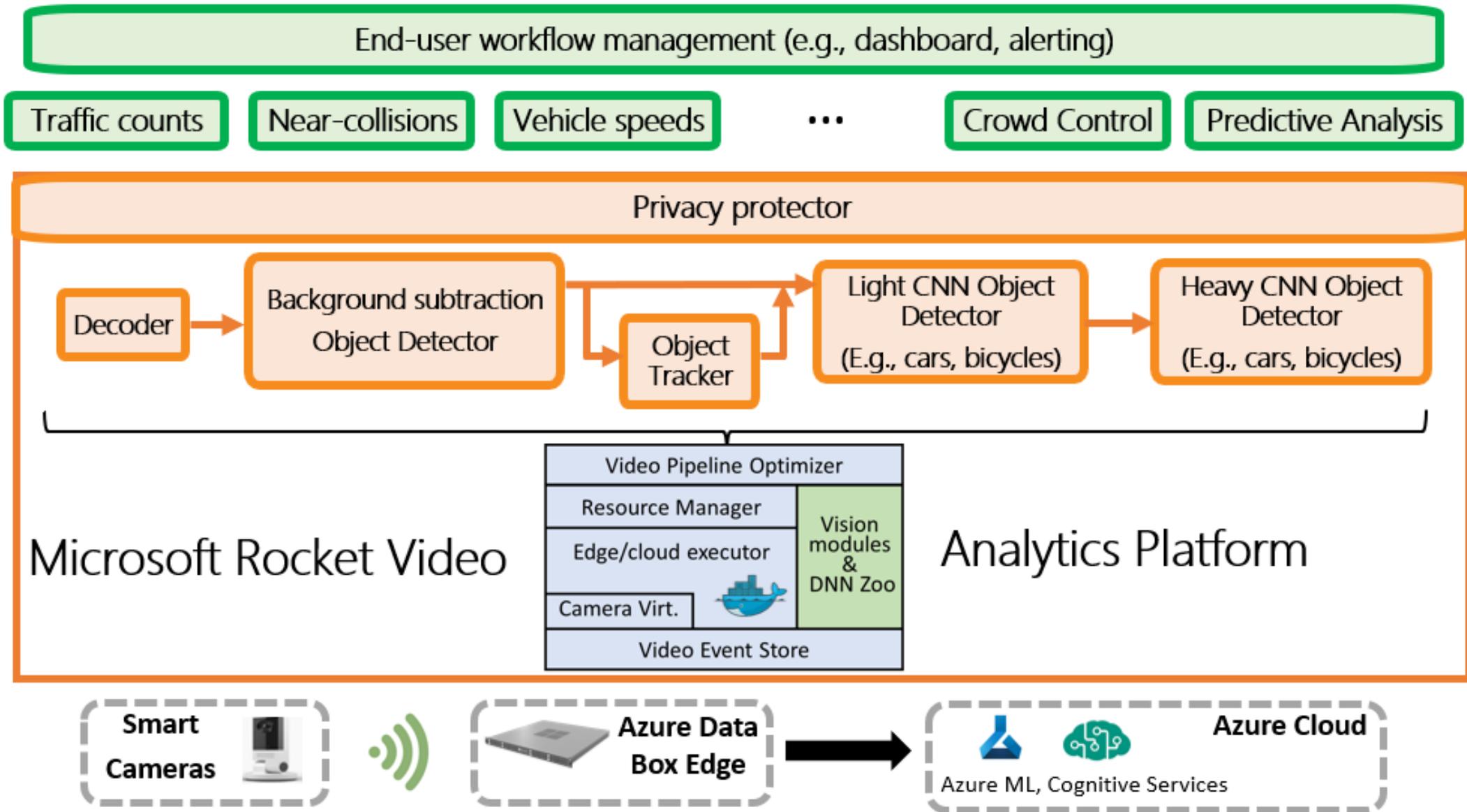
handling bandwidth sensitivity



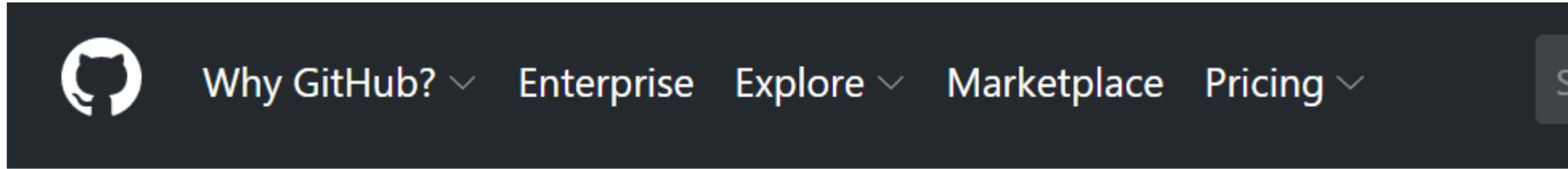








open source video analytics platform



microsoft / **Microsoft-Rocket-Video-Analytics-Platform**

<> Code

! Issues 0

🔗 Pull requests 2

▶ Actions

📊 Projects 0

🛡 Security

Project Rocket platform—designed for easy, customizable live video analytics—is open source

January 22, 2020 | By [Ganesh Ananthanarayanan](#), Principal Researcher; [Yuanchao Shu](#), Senior Researcher; [Landon Cox](#), Principal Researcher; [Victor Bahl](#), Technical Fellow, Director Mobility & Networking



nature electronics

Focus

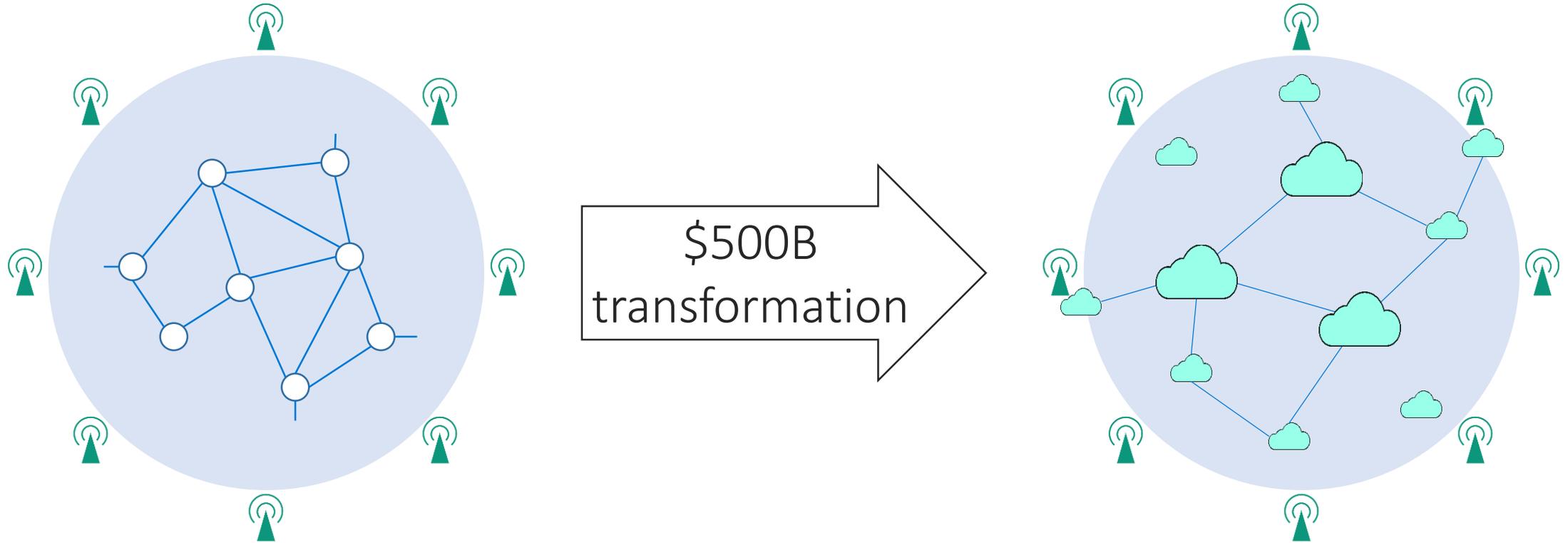
Technology of the Year 2019: Edge computing



what's in our future?



Telecom infrastructure transformation

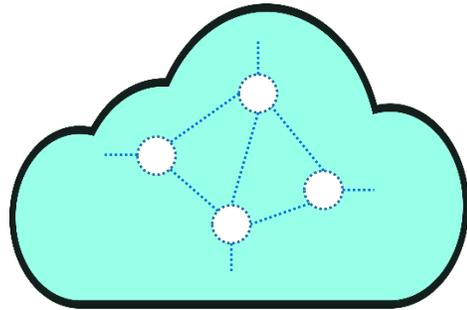


edge as communications infrastructure

Telecom network

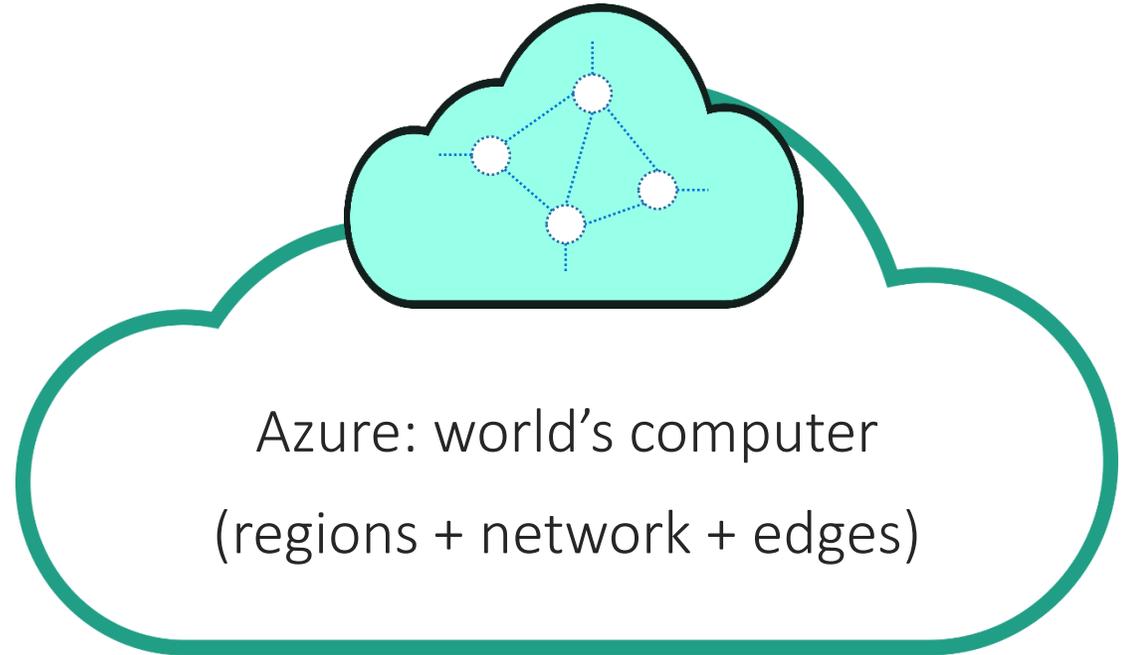
-- vs --

cloud providers



Open Stack, etc.

commodity
hardware



Azure: world's computer
(regions + network + edges)

AT&T, Microsoft Team Up In Massive Cloud, 5G Deal Reportedly Worth Billions

The multi-year deal, which is reportedly worth more than \$2 billion, will have AT&T and Microsoft collaborating on cloud and bringing new 5G and IoT solutions to the market.

By [Gina Narcisi](#)

July 17, 2019, 11:41 AM EDT

News

Nvidia Reveals Edge Supercomputer, Taps Ericsson for 5G vRAN



[Matt Kapko](#) | Editor

October 22, 2019 9:39 AM

Share this article:



Living On The Edge Will Make 5G Relevant



Frank Palermo Forbes Councils Member
Forbes Technology Council COUNCIL POST | Paid Program
Innovation

Nokia to provide 5G, edge computing, private LTE, drone systems in Sendai

Wednesday 18 October 2017 | 14:51 CET | News

Nokia and the City of Sendai have signed a MoU to deliver services for local businesses as they recover from the 2011 earthquake and tsunami. The agreement covers public safety management, including disaster recovery, activities to improve the safety and security of local

Intel is buying a Canadian software business for \$27 million to turbo-charge its edge computing and 5G efforts



Stephen Nellis, Reuters Oct 16, 2019, 2:13 AM



A Clear Vision for Edge Computing on Mobile Networks

Operators to uncover more than \$1 billion in new latency-sensitive revenue by 2025

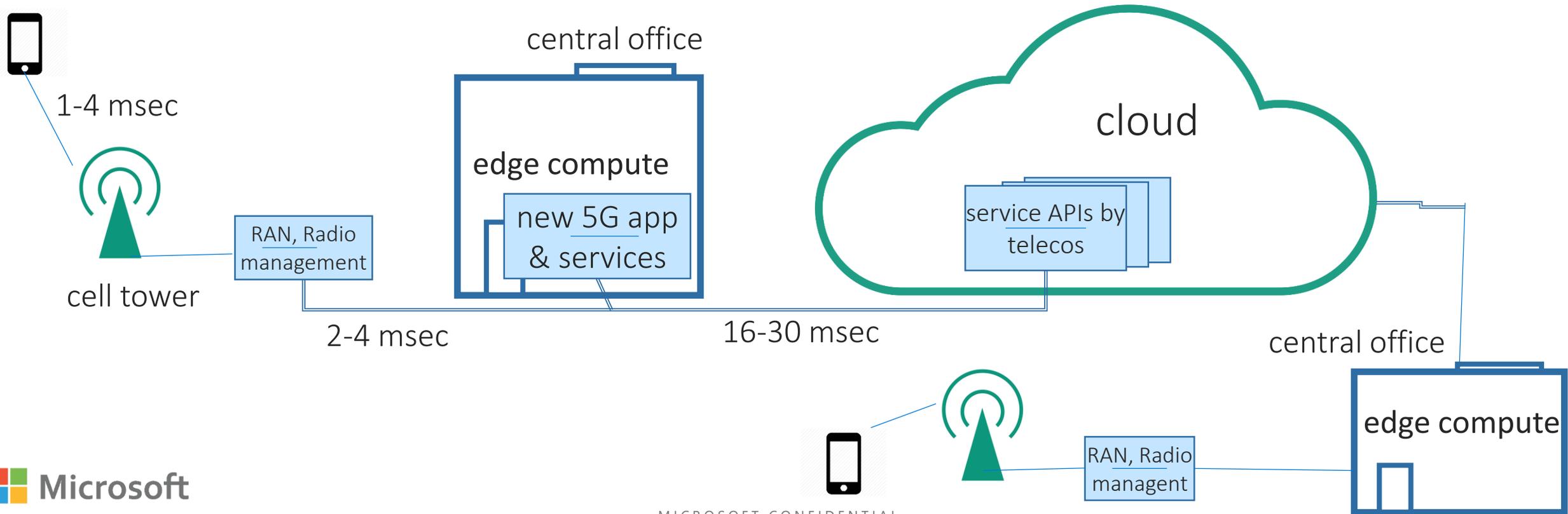
NEWS PROVIDED BY
[Mobile Experts](#) →
Apr 03, 2019, 08:33 ET

SHARE THIS ARTICLE



vision: a unified platform across edges, CO & cloud

- run any network functions anywhere (e.g. vRAN @ edge/CO, OTT in the cloud)
- IT companies join forces with telecoms to deliver the core infrastructure & service delivery platform
- Internet replaced by well managed, low latency IT WANs



win-win for both sides

telecom

- reduced operational burden / cost
- lower future infrastructure cost
- faster deployment of new features
- programmer eco system (no longer “just a pipe”)
- increase revenue from services



cloud companies

- additional revenue stream from selling network services & operations
- major boost to current bets on next gen apps (e.g. video analytics)
- better IoT story

thanks!



@SuperBahl