The Present and Future of Edge Computing From an International Perspective

Richard B. Dasher, Ph.D.

Director, US-Asia Technology Management Center

Adjuct Professor, Stanford University

Outline

- **About this series**

  - Casual attendance, credit requirements, video recordings

- **Edge computing**

  - Definition, history, basics of tech stack

  - Adoption

- **Drivers of edge computing adoption**

- **Upcoming sessions in the series**

  - And, what we could not include

Welcome to Everyone!
(Visitors & Registered Students)

♦ **Weekly <u>public</u> lecture / panel discussion series** presented by the US-Asia Technology Management Center

   ♦ This year: 27th year of this series

   ♦ Every Thursday (but Thanksgiving) from today through December 5, 2019

♦ **About the US-Asia Technology Management Center**

   ♦ Industry affiliates program (supported by membership and other fees paid by companies)

   ♦ For 25 years in EE, now under Stanford Global Studies; see http://asia.stanford.edu

♦ **Continuing mission:** Introduce trends and current developments at the intersection of a particular area of business and technology

   ♦ With special reference to U.S. – Asia: comparisons, cooperation, and competition

   ♦ Speaker slides and videos available for previous years at https://asia.stanford.edu/courses-events/public-lecture-series/

Registering for university credit for this series

- **Available to Stanford students**

- **EASTASN-402A "Topics in International Technology Management"**

  - Cross-listed as **EE-402A, EALC-402A**

  - No pre-requisites, open to undergrads and graduate students

  - May be repeated in future years for credit

- **One unit, pass / no credit (S / NC)**

  - "Seminar" course

  - Beware: some departments limit how many credits from "seminar courses" can apply toward graduation requirements

Seminars 402A – Requirements for Credit

See Syllabus for official statement of credit requirements

REQUIREMENTS MAY BE DIFFERENT THAN FOR OTHER SEMINARS

A. **In-person attendance** at all but two sessions (e.g. 8 of 10 sessions)
  - This Requirement (A) is waived for students registered through SCPD
  - Evidence of attendance is required: today fill out survey, from next week sign weekly pass-around sheet at auditorium – no signature, no credit!

B. **Submit one written comment / summary per session**
  each week for nine (9) of the ten (10) sessions
  - To me (Prof. Dasher) <rdasher at stanford dot edu>
  - cc to course assistant
    Sijia Mao <sijiamao at stanford dot edu>>

  - Comment must provide evidence that you watched the session
  - Each comment is due within two weeks of the date of the session
  - See Syllabus for details on formatting, etc. (**no attached files**)
 - Comments for today are due by October 10, 2019

Video recordings

- **Video cameras located in the back of the room will capture the presentations in this course**
    - Videos will be posted to Canvas and ultimately to Stanford's YouTube channel – freely available to the public
    - US-ATMC provides links to videos and speaker slides of this series (and previous series) on our website http://asia.stanford.edu
- **Video recording is an important aspect of this series**
    - Resource for students in fulfilling the credit requirements
    - Part of the core mission of the US-ATMC: improve knowledge-sharing between technical business communities in Asia and Silicon Valley
- **Cameras are primarily recording the lecturers & instructor**
    - Occasionally images of people in audience are incidentally captured.
    - Before video is made public, editors review the recordings and blur student images.
    - Voices are captured during Q&A
- **Concerned? See Dasher or course assistant**

Edge computing:
The next major architecture of networked information processing

Major Stages of networked information processing

3. Cloud computing

4. Edge computing

2. Server – (PC) client

1. Mainframe – terminal
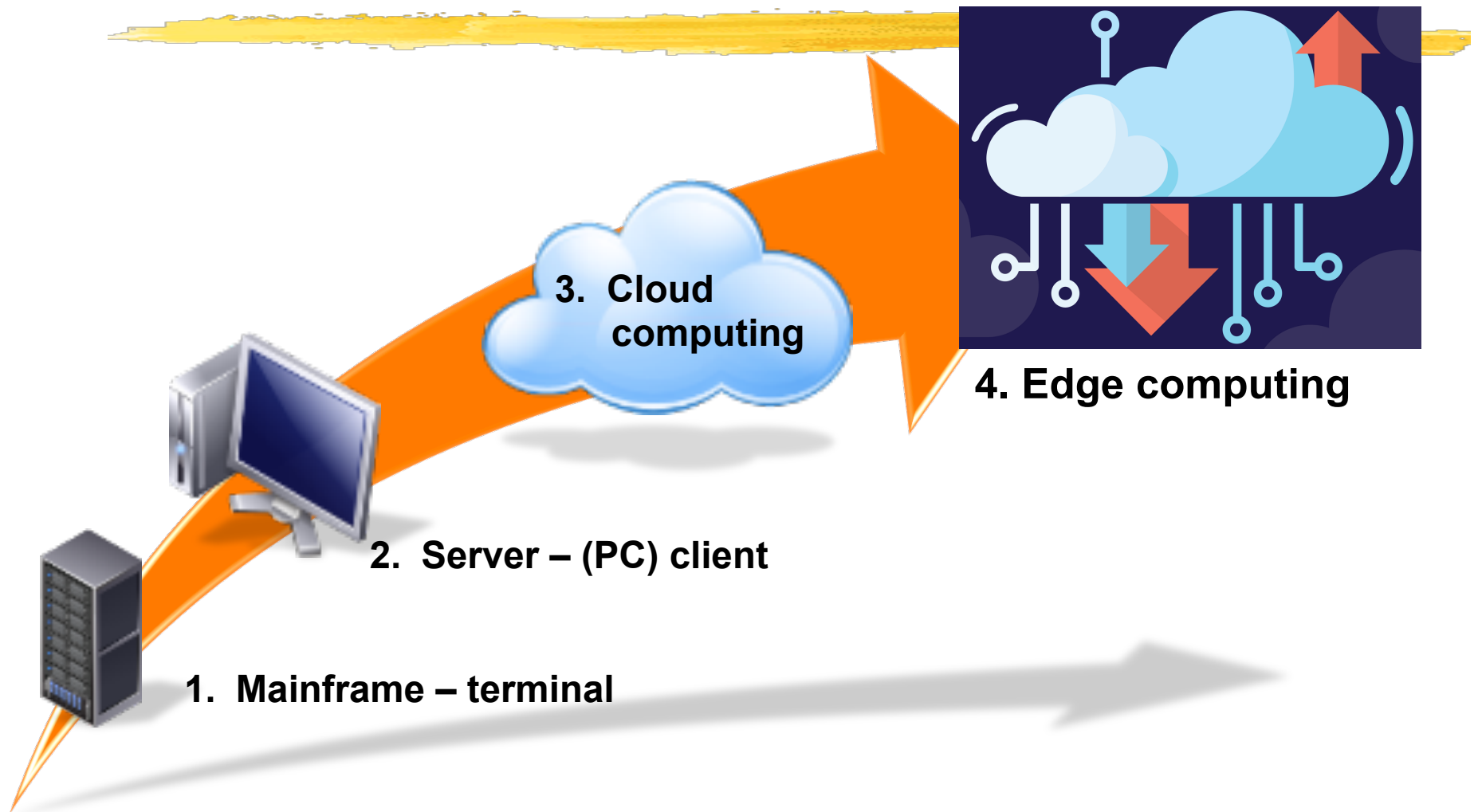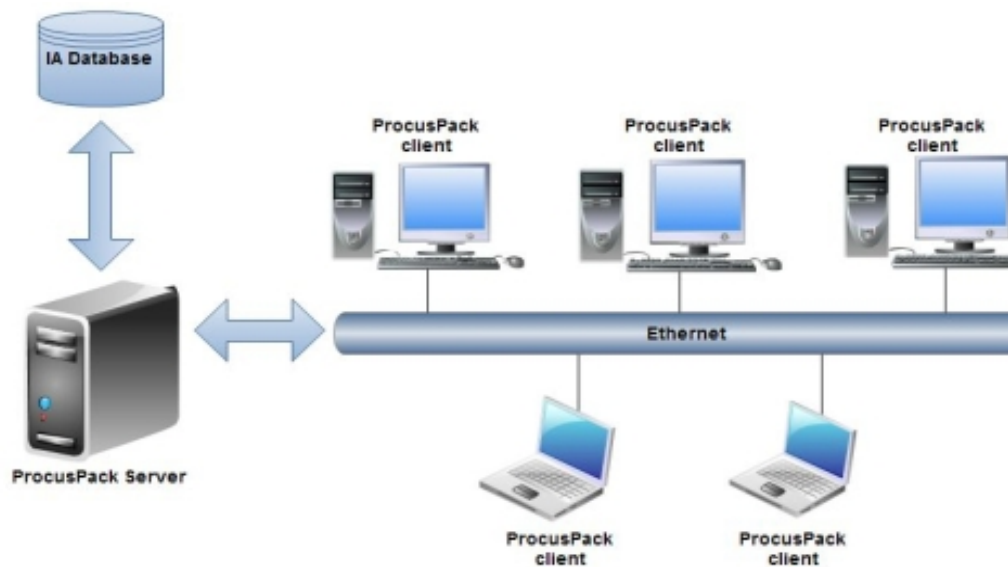
Mainframe – terminal (1940's – c. 1980)

| |
|---|
| **Clients**: (dumb) terminals – text only display |
| **Connectivity**: clients link to mainframe via private custom networks; time-share access |
| **Application programs**: custom-built for each system |

Client – server (c. 1980 – c. 2005)

IA Database

ProcusPack client · ProcusPack client · ProcusPack client

Ethernet

ProcusPack Server

ProcusPack client · ProcusPack client

Diagram from Procus (company) website

	Early Stage	Late Stage
	Clients: Mostly workstations	**Clients:** PCs, workstations, peripherals
	Ethernet LANs, spread of TCP/IP (Internet links)	**Internet everywhere** (LAN and open)
	Standardized applications run on the client, processing-intensive programs run on server	**Standardized "office" applications on the client;** other apps on server; may use central dbase

Cloud computing (c. 2005 – present)
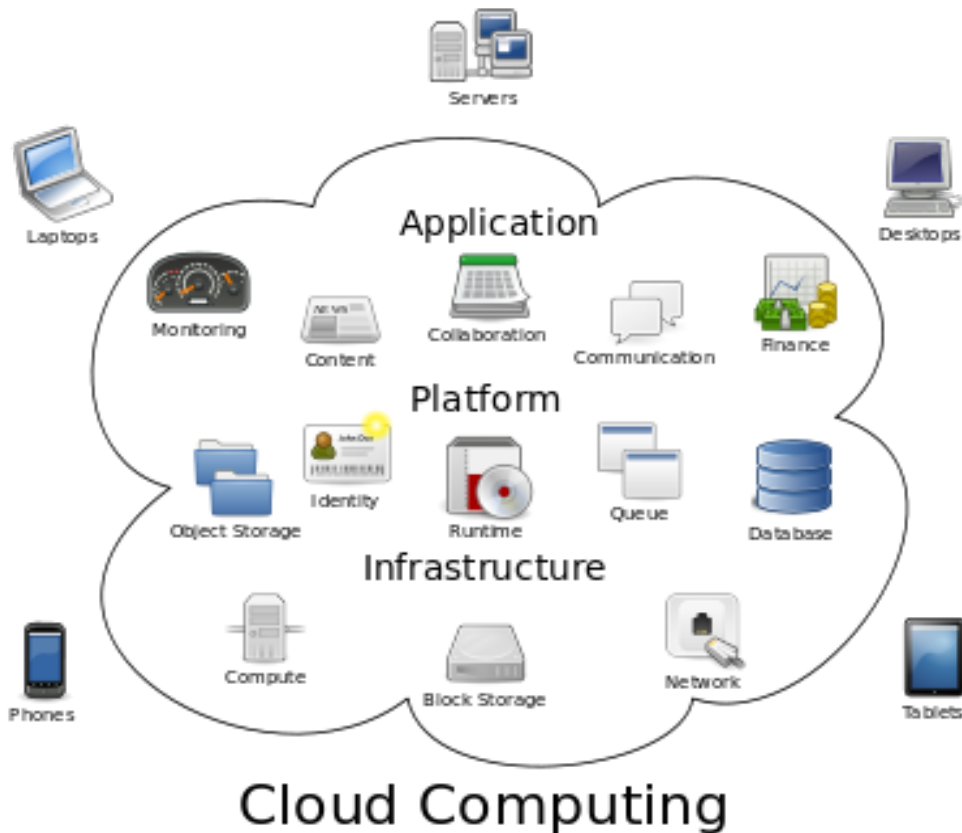
Diagram from Wikipedia

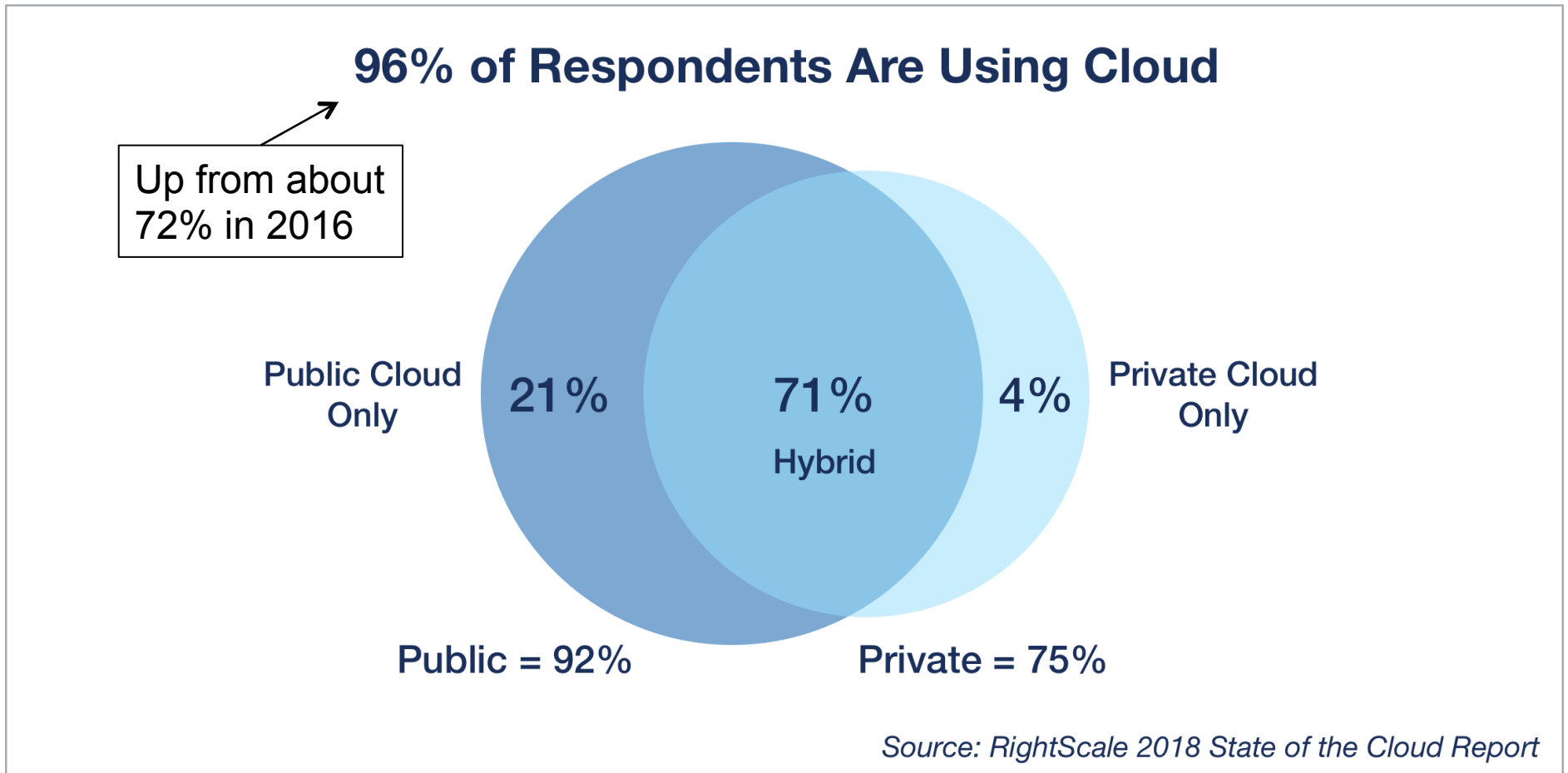**Many client types:** smartphones, PCs, tablets, MP3 players, sensors, smart appliances, autos …

Clients network to data centers & other clients via Internet; many LANs are just secure channels over public Internet

(Most) applications run on "virtual machines" in data center(s). Clients access via browsers; physical location of data may be distributed (even unknown)

By 2018: Cloud Computing is Ubiquitous

96% of Respondents Are Using Cloud

Up from about 72% in 2016

Public Cloud Only

21%

71%
Hybrid

4%

Private Cloud Only

Public = 92%

Private = 75%

Source: RightScale 2018 State of the Cloud Report

(January 2018 survey of 997 IT professionals, of whom 53% represented organizations of 1,000 or more employees)

2018 Cloud apps

Small company = < 100 employees, medium = 100 – 499, large = 500+

Cloud Application Usage	Small	Medium	Large
Email	72%	75%	73%
Web presence	72%	73%	72%
Business productivity	67%	75%	77%
Collaboration	62%	77%	76%
Virtual desktop	65%	75%	68%
Financial management	61%	75%	66%
Analytics	62%	72%	74%
CRM	54%	73%	68%
VoIP	61%	67%	62%
HR management	56%	70%	65%
Help desk	62%	66%	56%
Expense management	56%	71%	60%
ERP	52%	69%	62%
Call center	50%	55%	57%

Edge computing

Intelligent clients "at edge, e.g. new smartphones, connected cars, airplanes (UAVs & drones), robots, etc. **plus edge gateway servers (for less intelligent clients)**, e.g. sensors in a smart building, factory …

Integration of very high speed mobile and wired connectivity, e.g. 5G networks

Optimization of information processing between cloud (data centers) and edge, e.g. self-driving car operations at edge, biz analytics in cloud

Some important points about edge computing

♦ **Edge computing does not replace cloud computing**

 ♦ Much information processing will continue to be done in "cloud" data centers

 ♦ Processing at edge will focus on functions in which it's important to avoid latency (delays) or to avoid data transfer for other reasons

 ♦ Processing (e.g. business analytics) that require large amounts or heterogeneous data will stay in the cloud

You will also hear the term "fog computing"

 • a decentralized architecture in which data, compute, storage and applications are located somewhere between the data source and the cloud"

Cloud

Fog

Devices
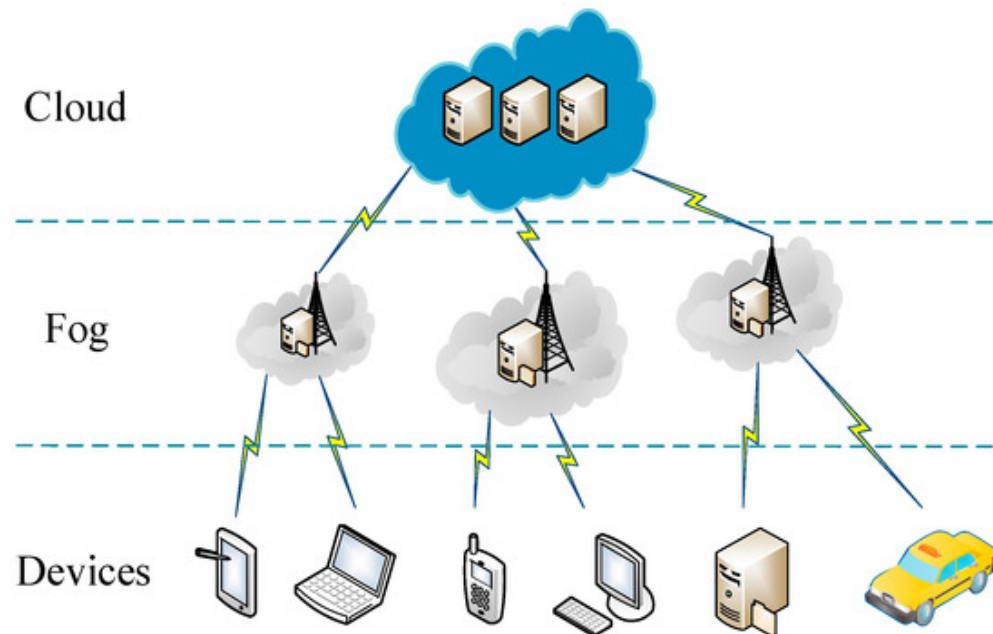
What is driving the advent of edge computing?

♦ **Huge increase in total data generated**
  - ♦ Especially related to advent of "Internet of Things" (IOT)
  - ♦ Much of the data that will be created is transient – why bother to clutter up the "cloud" data centers?

♦ **Huge increase in demand for real-time data processing**
  - ♦ No one wants a self-driving car to have to send data to the cloud and back in order to avoid an obstacle

♦ **Concerns about data security and privacy**
  - ♦ Data processing usually involves copying data from where it is stored to a "database" in order to perform search and other functions on it
  - ♦ E.g. analyze medical data in the hospital data center, and do not copy it into the cloud
  - ♦ Note: IOT security (at the edge devices or local servers) an issue

♦ **Accelerated computer chips for AI processing**
  - ♦ Application-specific chips (ASICs) can be much, much faster than multifunction processing chips (in data centers) and take less energy

New types of data coming into the cloud ...

New types of data are being <u>added</u> to legacy data

Structured text and numbers
- POS
- Financials
- Online forms (HR, travel, etc.)
- SQL databases
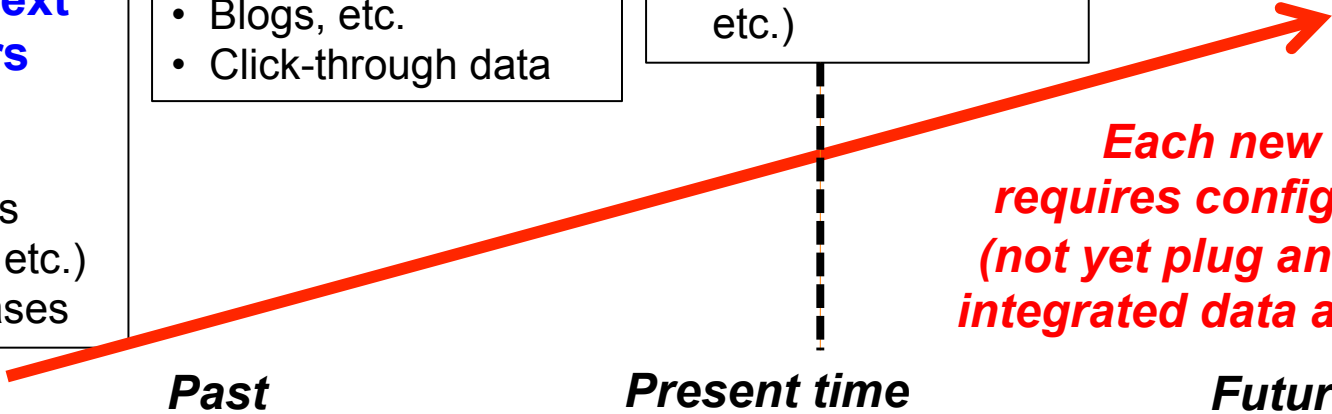
Unstructured text and numbers
- Social media
- Blogs, etc.
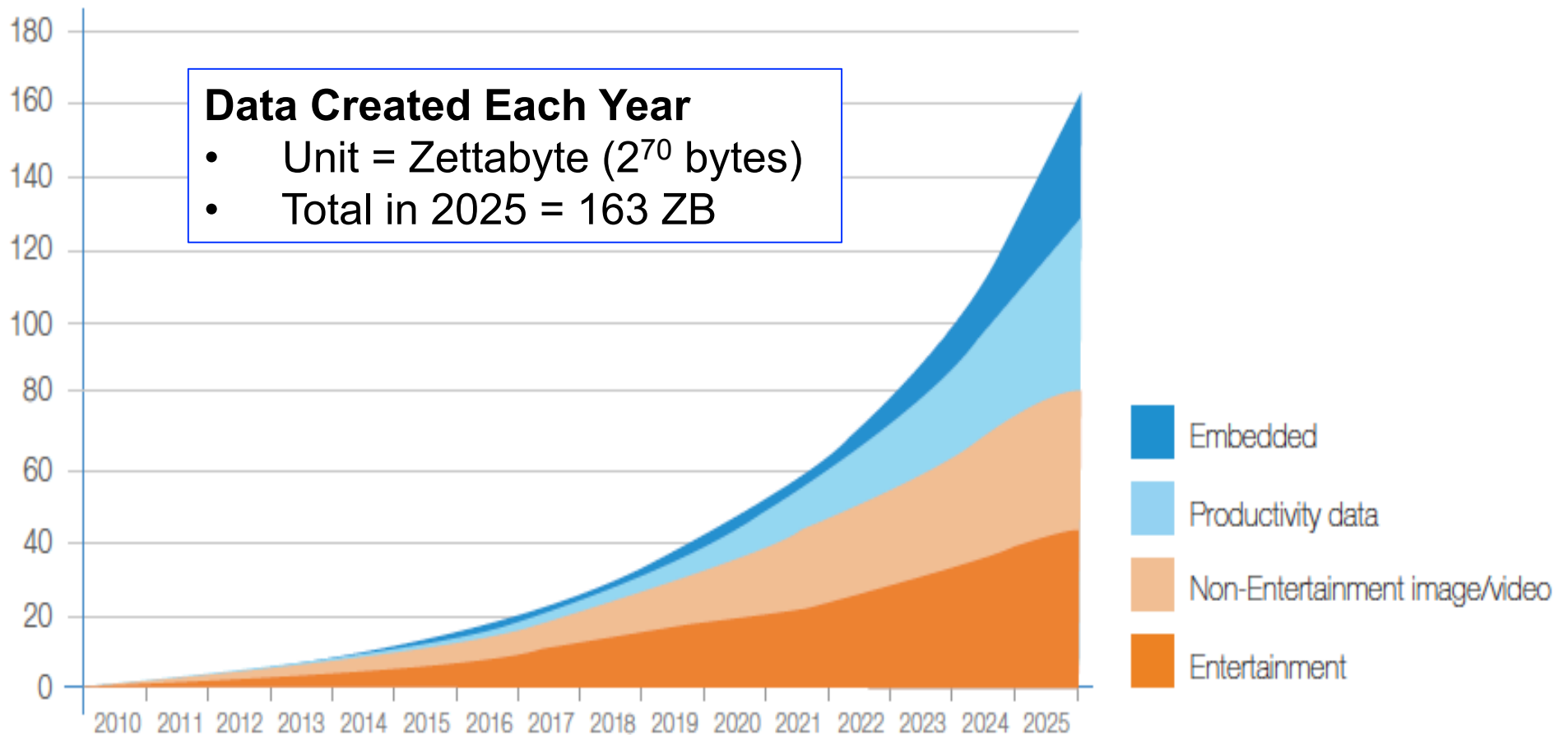- Click-through data

Images and audio
- Product ID and comparisons
- Intelligent assistants (Siri)
- TV and radio (inc. second screen, etc.)

IOT
- Sensor data
- OT – IT integration
- Connected smart devices
- Robots in cloud-based systems

Each new system requires configuration (not yet plug and play), integrated data analysis

Past   *Present time*   *Future*

... cause drastic increases in amount of data created

Data Created Each Year
- Unit = Zettabyte (2^{70} bytes)
- Total in 2025 = 163 ZB

Legend:
- Embedded
- Productivity data
- Non-Entertainment image/video
- Entertainment

IDC Survey. April 2017. *Data Age 2025 (Figure 6)*, sponsored by Seagate
<https://www.seagate.com/files/www-content/our-story/trends/files/Seagate-WP-DataAge2025-March-2017.pdf>

Real-time data (logs, locations, sensor output, etc...) is becoming a bigger share of all data

Real-time data
- = almost 30% of all data created (2025)
- Almost all real-time data will be IOT data (this chart)
- Much real-time data will come in "spikes" − wide variation in loads

Source: IDC's Data Age 2025 study, sponsored by Seagate, April 2017

IoT   Other
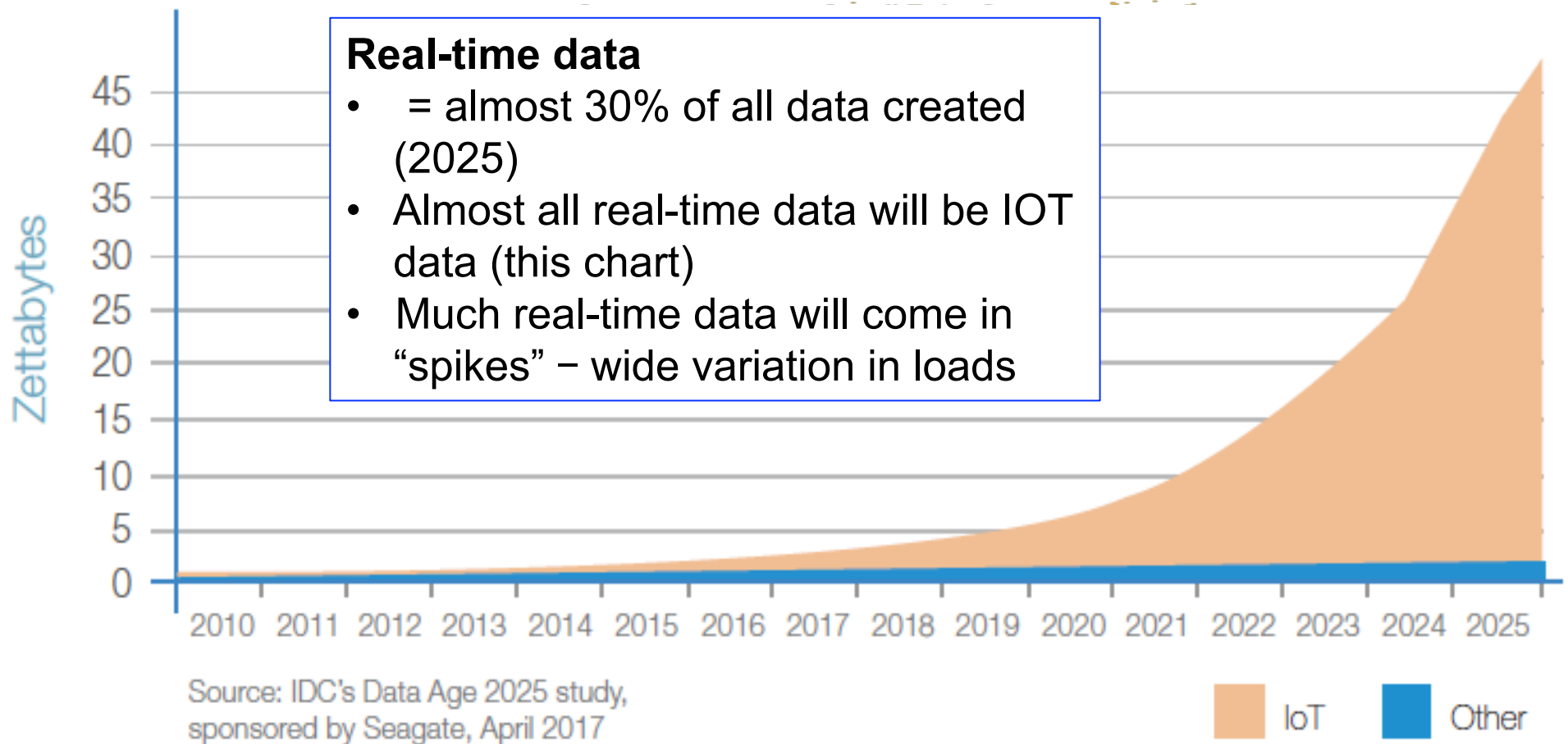
Market predictions for edge computing
(market research reports published during 2019)

$ billions

Edge computing	2017 - 2018	2019		2024	2025	CAGR (%)
Firm A	$1.47				$28.84	54.0
Firm B	1.73				16.56	32.8
Firm C		2.8		9.0		26.5
Firm D	1.2			6.96		34.0
Firm E	2.17			9.22		27.3
Cloud computing						
Firm X					712.83	18.5
Firm Y	36.7				285.30	29.2
Firm Z		319.8			696.25	10.2

On the market impact of edge computing (comments on previous slide)

- **Cloud computing markets predicted to stay much bigger than edge computing markets**

- **Edge computing markets predicted to have CAGR than are higher than cloud computing in general**

- **Disagreements in calculating market size and growth probably result mostly from differences in definition**

  - Narrow definition: only edge computing services

  - Broader definition would include larger ecosystem for edge computing

    - Some portion of the growth of related technology businesses (e.g. 5G networks)

    - Some portion of use-case spending on applications that utilize edge computing architecture (connected cars, energy systems, etc.)

Some use cases for edge computing

♦ **Smart buildings** (including homes, factories, warehouses, ...)

♦ **Smart manufacturing** (process control, Industry 4.0)

♦ **Self-driving cars, drones, airplane systems, etc.**

♦ **Connected car services**

♦ **Mobile augmented reality (AR), networked gaming**

♦ **Mobile ecommerce** (register-less stores, mobile banking, etc.)

♦ **Healthcare monitoring** (wearable devices, robot assisted surgery, medical data analysis in the hospital)

♦ **Physical security** and surveillance systems

♦ **Energy systems** (smart grids, etc.)

♦ **Some tele-medicine, agricultural systems control**, etc.

♦ **Virtual assistants** (Alexa) – maybe

Edge computing: regional perspectives

Edge computing is major topic of interest in Asia

♦ **Asia predicted to have the highest CAGR for edge computing markets (one firm predicts 40.5% APAC vs. 32.6% global)**

♦ **High rate of people in Asia say edge will be important**

  ♦ Jan 2017 online survey by VertivCo of 8,500 executives, IT professionals, other business people in SE Asia, N. Asia, Australia, New Zealand

  ♦ Question: Do you have plans to deploy edge initiatives in next 1 – 3 years?

  62%   38%
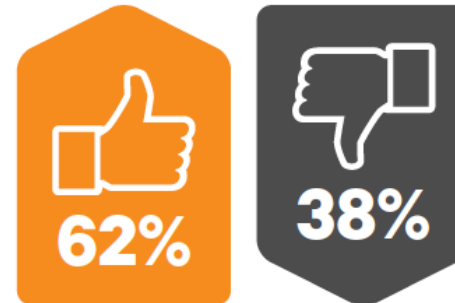
  ♦ Question: Do you think that the edge will be a relevant part of your business and IT strategy over the medium to long run (beyond 3 years)

  97%   3%
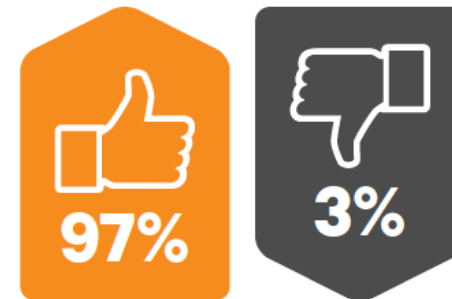
Will Asia move ahead of the U.S. in edge computing?

♦ **Asia has been deploying many projects related to edge computing − relatively independently of U.S.**

  ♦ China Unicom: virtual edge-cloud testbed in Tianjin City for video optimization and security monitoring

  ♦ China Telecom: "China Mobile Smart Parking" project in Yunnan and SE Guizhou (similar project by Chunghwa Telecom in Taipei)

  ♦ Baidu has software platform "Apollo" for autonomous vehicles

  ♦ Taiwan: smart streetlight system in Taoyuan City

  ♦ South Korea: world's first 5G enabled self-driving car test zone in Seoul (announced 3/2019)

  ♦ S. Korea to build nationwide 5G network in 2020

  ♦ Japanese companies NTT DoCoMo, KDDI, SoftBank, and Rakuten to invest $14.4 billion in 5G networks

  ♦ NTT Docomo (Japan): proof-of-concept project enables video analytics to be processed on surveillance cameras using MEC ("multi-access" or "mobile" edge computing)

Different technology directions for Asia and U.S.?

♦ **"Decoupling" of U.S. and China**

  ♦ Restrictions on U.S. firms from buying equipment from or selling hardware or software & services to Huawei, ZTE (currently under "temporary" easing

    ♦ Huawei developed own operating system (since cannot rely on access to Google Android)

    ♦ Huawei, Alibaba have announced their own AI processing computer chips

  ♦ Pressure by administration on U.S. firms to move manufacturing (especially final assembly) to U.S.

    ♦ But some are moving to SE Asian countries instead

  ♦ Tighter monitoring, control of foreign investors in U.S.

    ♦ Sinovation (Kai Fu Lee VC firm) has pulled back from U.S., focuses on China

    ♦ Chinese acquisition of AppLovin (AdTech) blocked; Chinese investor(s) ordered to divest from PatientsLikeMe (HealthTech)

    ♦ China FDI to U.S.: $46 billion (2016) – $27 bn (2017) – $4.8 bn (2018)

  ♦ China: new cybersecurity policies (coming online this autumn) appear to target foreign firms there

China regional influence increasing

- **China – IndoPac region trade (2018) = $2.5 T**
  US – IndoPac trade = $1.4 T
  - China accounts for 24% of IndoPac exports
  - U.S. accounts for 12% of IndoPac exports
- **RCEP trade area negotiations moving forward**
  - China-led alternative to TPP, includes 16 countries
- **Investments in SE Asia**
  - Some prominent: $2 billion by DiDi Chuxing and Softbank in Grab (2017)
  - Chinese VC investments in ASEAN startups increased 4x in 2019H1 to US$667M
  - Recent upswing in Belt and Road investments in SEA
    - $11 billion in 2019H1, $5.6 billion in 2018H2
  - Overall, Japan probably still has much bigger influence in SEA economies

How will regional divergence play out with regard to edge computing?

Topic of this series

Upcoming sessions in this series

♦ **October 3** **"The Promise of 5G"**

Prof. Arogyaswami Paulraj

- ♦ Critical enabling technology that will have very close relationship to continued growth of edge computing
- ♦ One of the key technologies in U.S. – China competition
- ♦ Prof. Paulraj won the Marconi Prize, his work has been critically important to WiFi and 4G networks
- ♦ World expert on 5G; frequent consultant & advisor

♦ **October 10** **"New chip technologies for AI and for Asia"**

Gary Brown, Director of AI Marketing, Intel

- ♦ New types of chips for faster AI processing with lower power consumption enable more edge computing
- ♦ AI chips causing a resurgence of chip industry growth
- ♦ Gary was with startup Movidius (acquired by Intel); former US-ATMC research assistant with extensive Japan & Asia experience

Upcoming sessions, p. 2

♦ **October 17 "Edge computing in autonomous vehicles"**
Dr. Sven Beiker, Dr. Maarten Sierhuis

‣ One of the most-often cited use cases for edge computing

‣ Panel discussion with former exec. director of Stanford CARS Lab and CTO of Alliance Innovation Lab SV (Nissan research group)

♦ **October 24 "Federated learning in medicine"**
Dr. Thomas Clozel, Founder / CEO, OWKIN

♦ "Federated learning" allows analysis of edge data without copying them to a data center

♦ An application of edge computing for security, privacy

Upcoming sessions, p. 3

- **October 31 "tinyML – enormous opportunities ahead"**
  Dr. Evgeni Gousev, Senior Director, Qualcomm
  Mr. Pete Warden, Research Engineer, Google
  - A new industry-led consortium with a solution for AI processing on edge devices
  - Evgeni has mentored much Ph.D. research at Stanford; Pete is a well-known blogger

- **Nov. 7   "Edge computing and the evolution of AR / VR"**
  Mr. Dijam Panigrahi, COO & Co-Founder, GridRaster, Inc.
  - Edge computing is important to improving user experience with augmented reality or virtual reality (latency may make users may feel ill)
  - GridRaster provides a collaborative mobile device – edge cloud platform for mixed-reality experiences with mobile devices

- **Nov. 14   "Japanese startup use cases of edge computing"**
  Dr. Atsunori Kanemaru, Chief Scientist, LeapMind
  - LeapMind has delivered edge computing POCs to major firms (e.g. subway train door management)

Upcoming sessions, p. 4

◆ **Nov. 21**   **Speaker still TBD**

◆ **Nov. 28**   **Thanksgiving – no class**

◆ **Dec. 05**   **"Possibilities for edge computing in U.S. and Asia" (title tentative)**
Dr. Yoky Matsuoka, Vice President, Google

◆ Robotics expert

◆ Former professor at CMU

◆ Former CTO of NEST

◆ "Fireside chat" will include discussion of challenges posed by edge computing, international trends in development