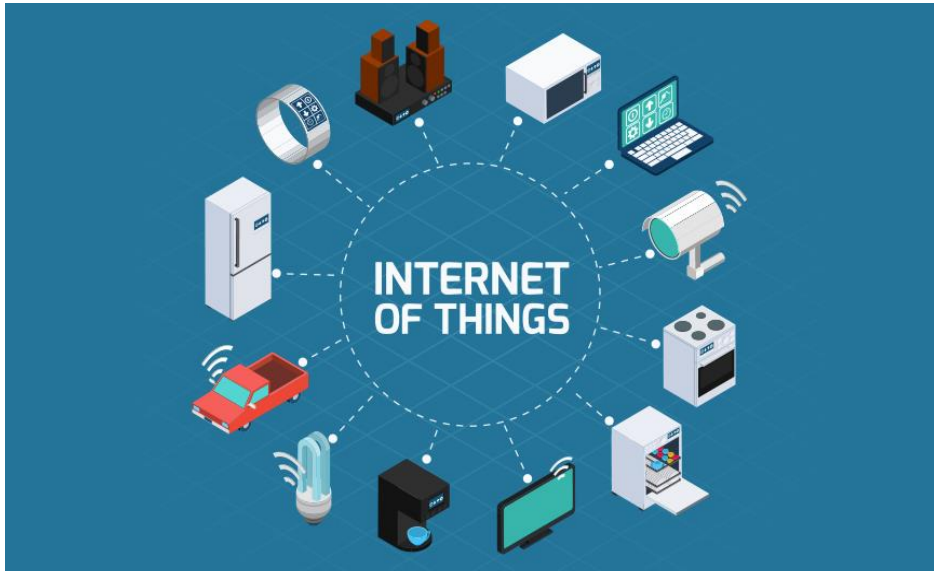
Internet of Things: Protocols and Networks (CSC2106)

Cloud Computing

Cloud vs Edge Computing

Traditional IoTs: Relies on Cloud Computing

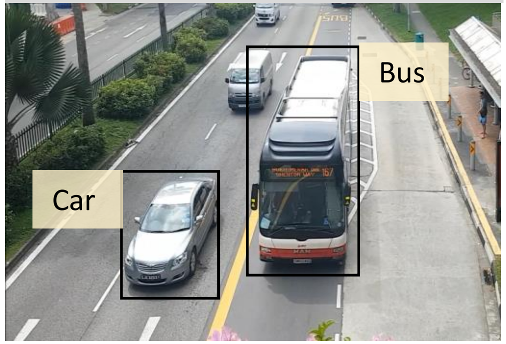


Source: https://www.bankinfosecurity.com/gao-assesses-iot-cybersecurity-other-risks-a-9926

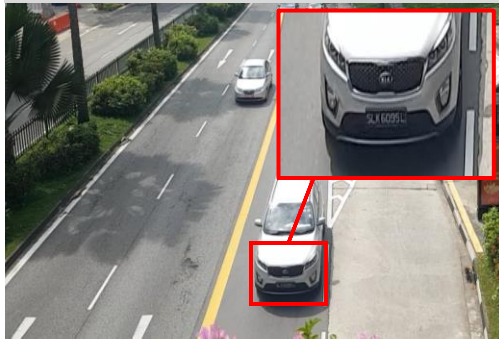
Modern IoTs with Edge Computing

Crowd Counting

Object Localization & Classification

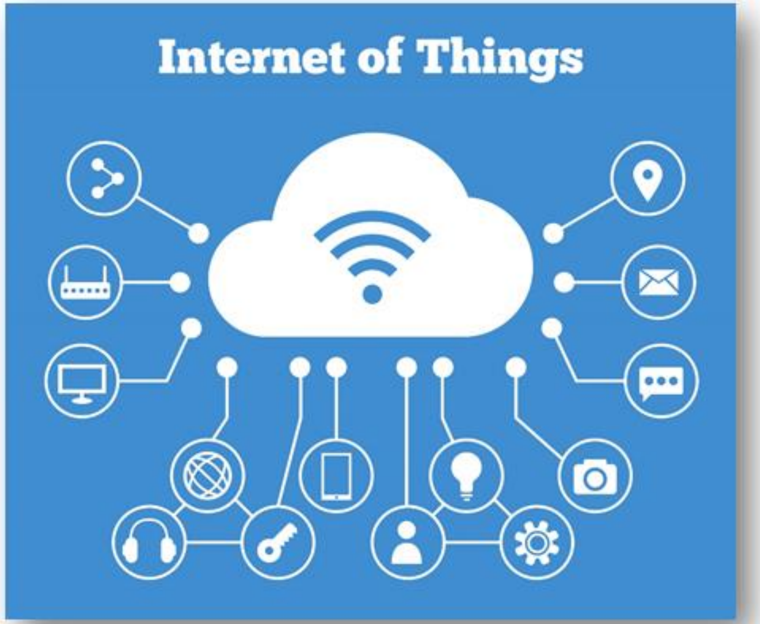


License Pate Recognition



Cloud Computing

Cloud for general computing



Cloud for IoTs

Primary difference in number of connected devices

Genesis of Cloud Computing

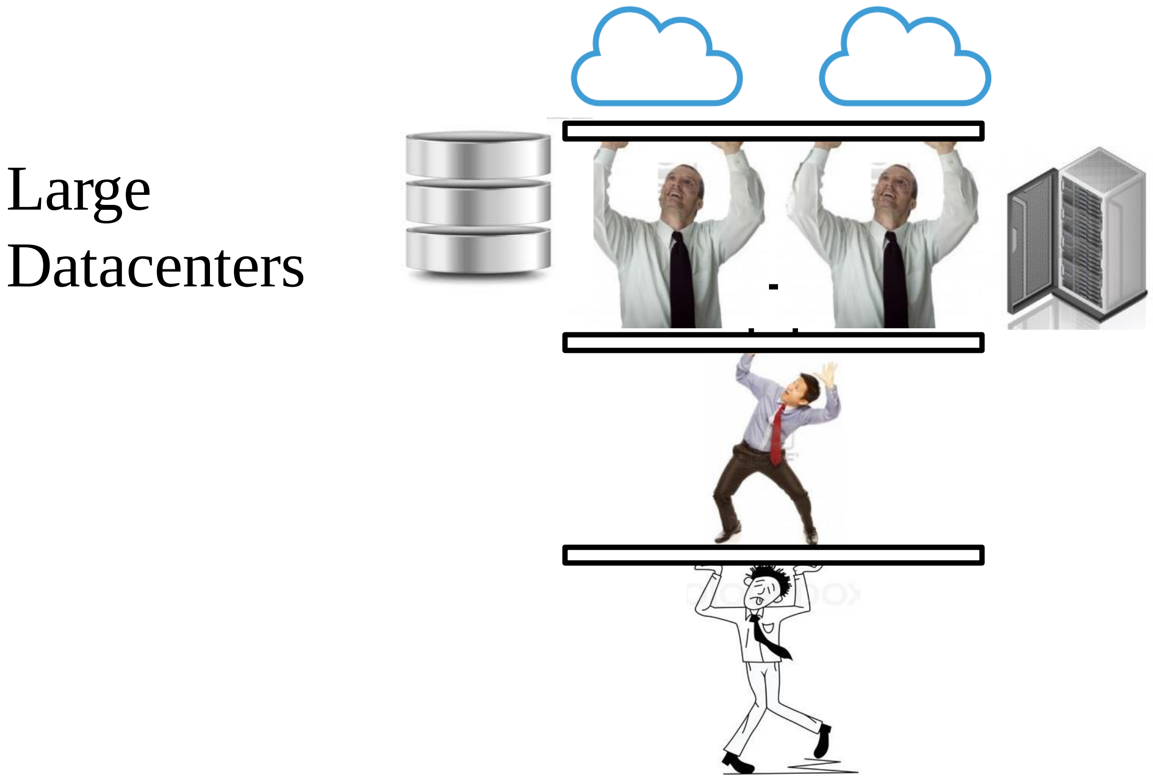
- August 25, 2006: Amazon announced EC2 (https://aws.amazon.com/blogs/aws/amazon_ec2_beta/)
  - Birth of Cloud Computing in reality

Unverified Folklore about the birth of EC2

- Amazon's CEO was amazed by the number of computers in their datacenter and their low utilization
- Computing facilities are designed for peak load (Christmas)
- Needed a way to rent unused capacity, like renting their warehouses and other infrastructure
  - Develop an application programming interfaces (APIs) to remotely use computers.
- So began the computer rental business that we now call cloud computing.
- Sharing an underutilized resource is good for cloud service customers as well as for the cloud service providers.

Ref: Raj Jain and Subharthi Paul, "**Network Virtualization and Software Defined Networking for Cloud Computing - A Survey**," IEEE Communications Magazine, Nov 2013, pp. 24-31, ISSN: 01636804, DOI: 10.1109/MCOM.2013.6658648,

Cloud Enabled by Networking



- Clouds
- Low Average Demand
- Virtualization Storage/Compute/ Network
- High Speed Networking

Ref: http://www.cse.wustl.edu/~jain/cse570-18/

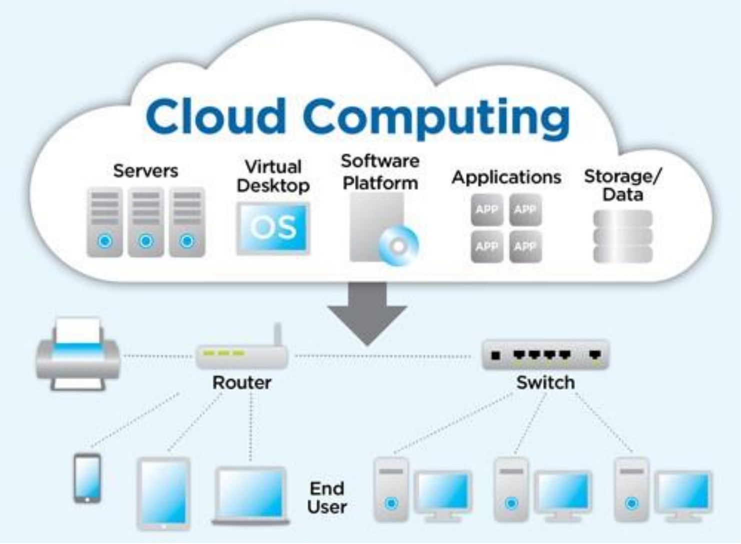
What is Cloud Computing?

Cloud computing is a model for enabling convenient, on- demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.

-National Institute of Standards and Technology (NIST)

http://www.nist.gov/itl/cloud/

Simpler definition: network-based computing taking place over the Internet, while hiding complexity of underlying infrastructure using simple APIs



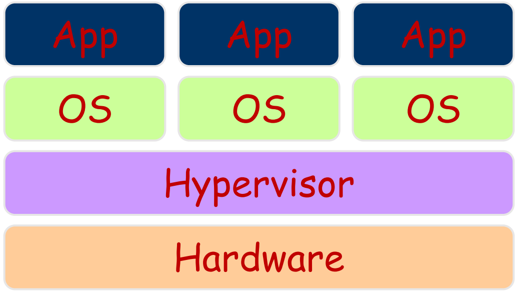
- Key Characteristics:
  - On-Demand self-service
  - Scalable
  - Shared
  - Ubiquitous access
  - Rapid provisioning/released
  - Minimal management

Advantages of Cloud Computing

- New applications
- Anytime/anywhere access
- Homogeneity
- Virtualization
- Resilient
- Cost
- Sharing, collaboration
- Management/maintenance
- Security
- …

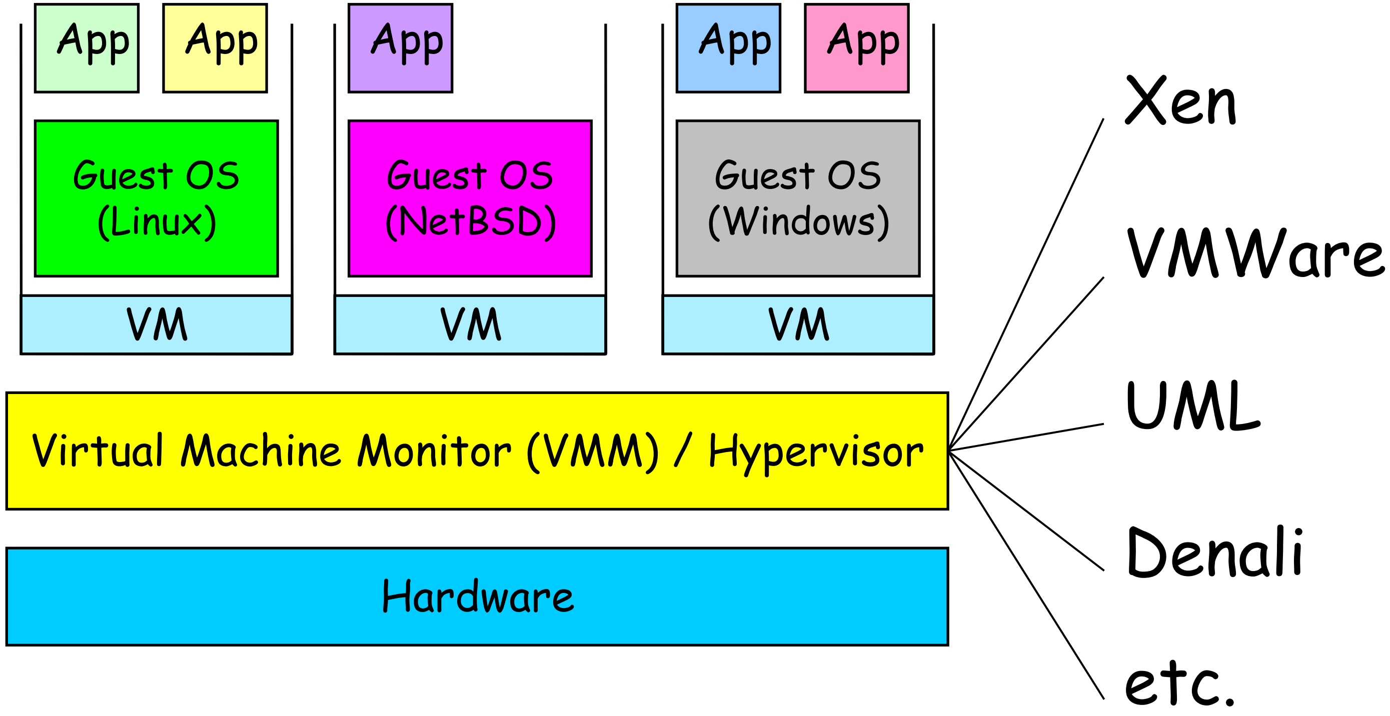
Virtualization

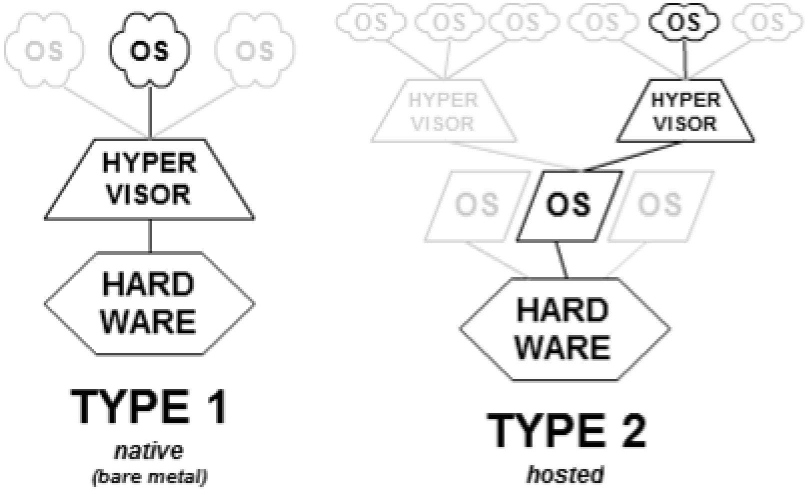
- Virtual workspaces:
  - An abstraction of an execution environment that can be made dynamically available to authorized clients by using well-defined protocols
  - Resource quota (e.g., CPU, memory share)
  - Software configuration (e.g., OS, provided services)
- Implemented on Virtual Machines (VMs):
  - Abstraction of a physical host machine
  - Hypervisor intercepts and emulates instructions from VMs, and allows management of VMs
  - VMWare, Xen, etc.



Virtualized Stack

Virtual Machines and Hypervisors



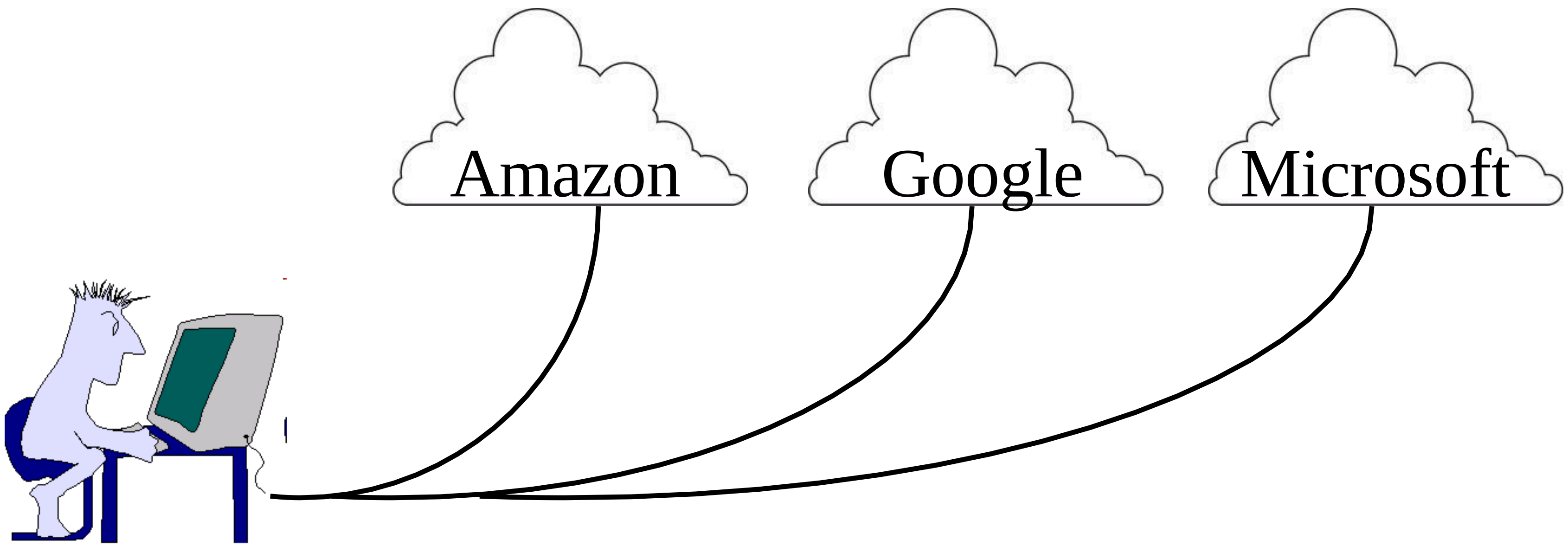


Two types of Hypervisors, Type 1 and Type 2

- Type 1: Bare metal hypervisor. e.g. VMware ESXi (formerly ESX) and Xen.
- Type 2: Runs on a host OS e.g. Parallels Desktop for Mac, QEMU, VirtualBox, VMware Player and VMware Workstation

Cloud Computing as a Utility

- Cloud computing is a utility like water, electricity, telephone
- Just turn on your computer and get computing storage and other resources as you need it and pay for it by usage.
- Computers include smart phones, tablets, laptops, desktops, …

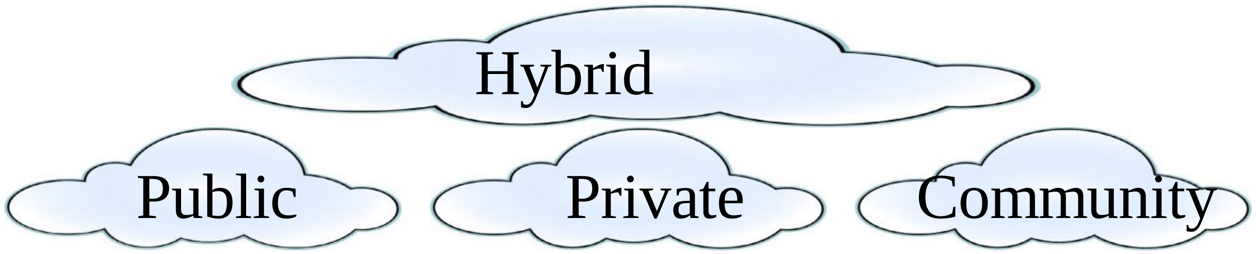


Barriers to Cloud Adoption

- Security:
  - Is our data secure?
  - How can we audit security?
  - Will my data be erased on deletion?
  - …
- Compliance:
  - Will we be compliant with risk, security, privacy laws?
- Interoperability:
  - Can I take my workload from provider to provider?
  - Not easily
- Service Level Management: Is the billing accurate? What happens on failures? Is the capacity sufficient?
- Tools:
  - How can I automate provisioning, monitoring, management

Ref: https://en.wikipedia.org/wiki/Cloud_computing_issues,
https://en.wikipedia.org/wiki/Cloud_computing_security

Cloud Deployment Models

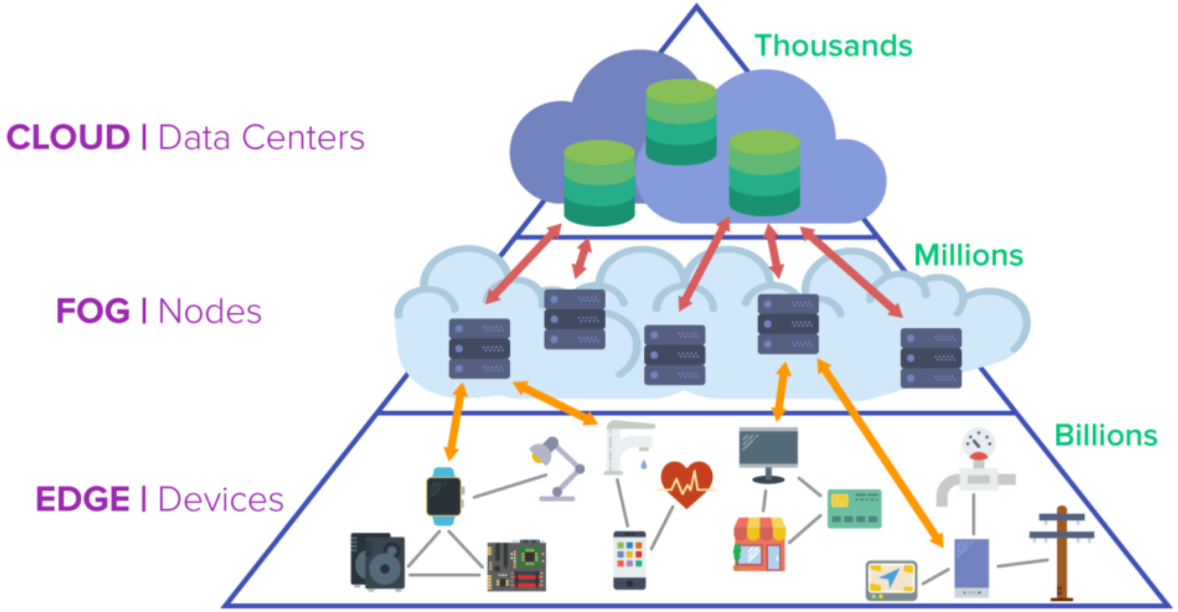


- Public Cloud: Offered by large companies for use by anyone who pays. Least expensive and least secure.
- Private Cloud: Owned by a specific entity and used only by that entity. More expensive and more secure.
- Community Cloud: Shared by multiple organizations who share a common cause. Did not materialize in practice.
- Hybrid Cloud: Cloud consisting of two or more Public, Private cloud. E.g., Private cloud for sensitive data and public cloud for less sensitive data
- Other examples such as Virtual Private Cloud, Multi-cloud, Distributed Cloud and Edge Cloud.

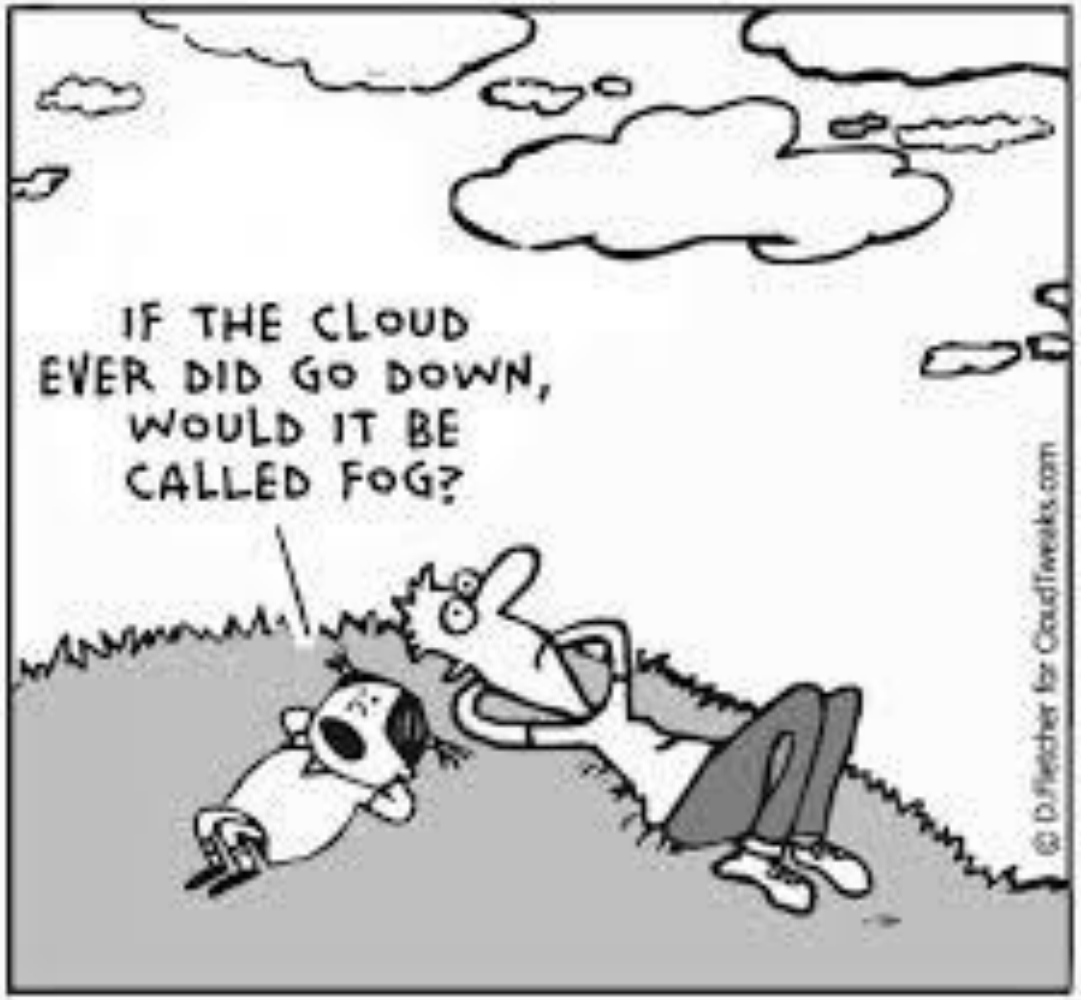
	Public	Private	Community	Hybrid
Ease of setup and use	Easy	Requires IT proficiency	Requires IT proficiency	Requires IT proficiency
Data security and privacy	Low	High	Comparatively high	High
Data control	Little to none	High	Comparatively high	Comparatively high
Reliability	Low	High	Comparatively high	High
Scalability and flexibility	High	High	Fixed capacity	High
Cost-effectiveness	The cheapest	Cost-intensive; the most expensive model	Cost is shared among community members	Cheaper than a private model but more costly than a public one
Demand for in-house hardware	No	Depends	Depends	Depends

Ref:https://www.sam-solutions.com/blog/four-best-cloud-deployment-models-you-need-to-know/

Fog Computing



Ref: https://www.omnisci.com/technical-glossary/fog-computing



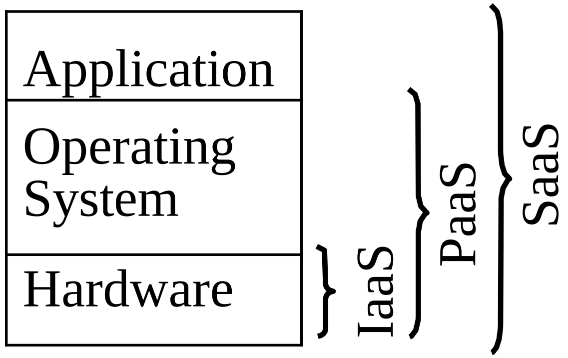
Ref: http://community.spiceworks.com/topic/254392-fog-computing-replaces-cloud-as-new-tech-buzzword

Fog Computing (Cont)

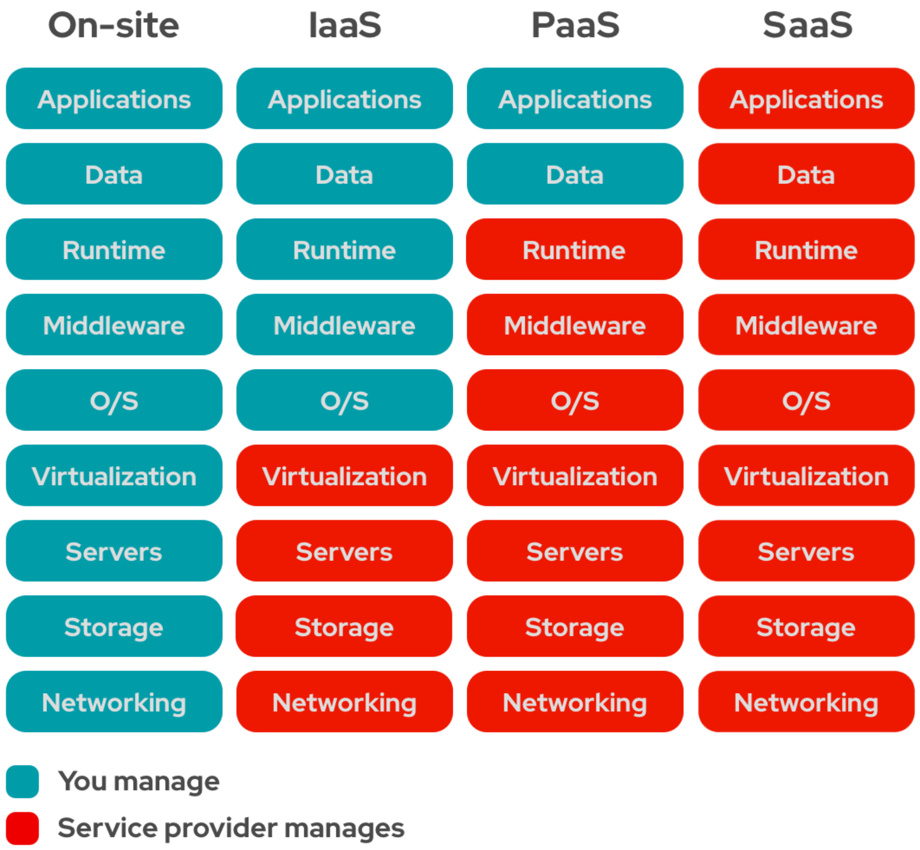
- Location Aware and Location Sensitive
  - → Low latency → Computing in micro clouds
  - → Computing in the edge → Computing everywhere
  - → Fog
- Geographically distributed → Everywhere/Anywhere
- Large Scale
- Mobility
- Real-Time

Ref: F. Bonomi, et al., "Fog Computing and Its Role in the Internet of Things," ACM MCC'12, August 17, 2012, Helsinki, Finland

Cloud Service Models

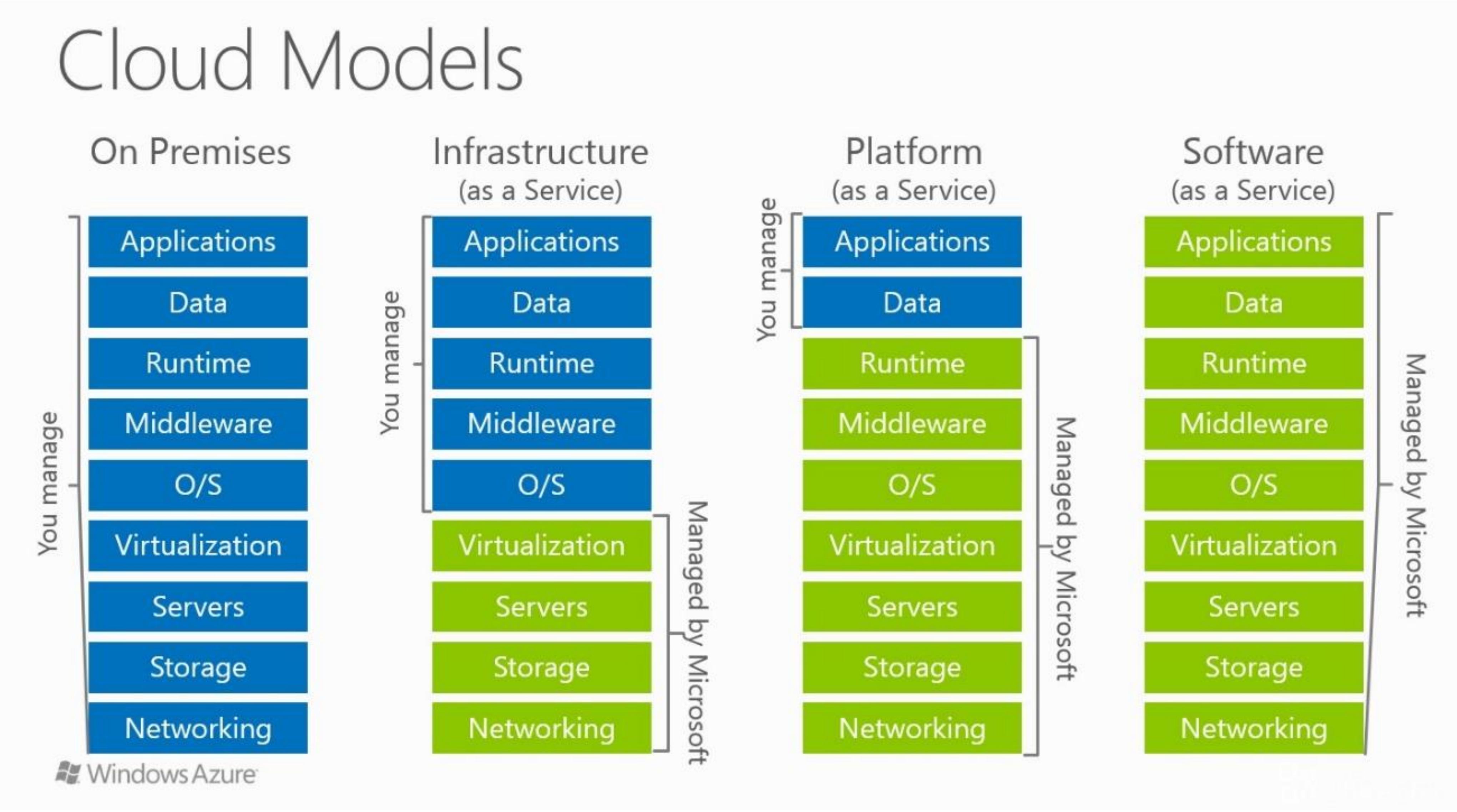


- **Infrastructure as a service (IaaS)**: Provide virtual machines, storage, and network facilities that users can manage. Users install their own operating systems and software, e.g., Amazon Web Services
- **Platform as a service (PaaS)**: Includes hardware and operating systems, development and administrative tools. Users develop and deploy their software, e.g., Microsoft Azure
- **Software as a service (SaaS)**: Complete application with a user interface, e.g., Google Docs, Gmail, Salesforce.com



Ref: https://www.redhat.com/en/topics/cloud-computing/iaas-vs-paas-vs-saas

Cloud Service Models example: Microsoft Azure



More examples of IaaS, PaaS, SaaS

- IaaS: Amazon Web Services (AWS), Microsoft Azure, Google Compute Engine
- PaaS: Google App Engine, Heroku, OpenShift, AWS Elastic Beanstalk
- SaaS: Google Apps, Dropbox, Cisco Webex, Salesforce, Concur, GoToMeeting

Everything As A Service

- https://en.wikipedia.org/wiki/Banking_as_a_Service
- https://en.wikipedia.org/wiki/Content_as_a_service
- https://en.wikipedia.org/wiki/Data_as_a_service
- https://en.wikipedia.org/wiki/Function_as_a_service
- https://en.wikipedia.org/wiki/IT_as_a_service
- https://en.wikipedia.org/wiki/Location_as_a_service
- https://en.wikipedia.org/wiki/Mobile_backend_as_a_service
- https://en.wikipedia.org/wiki/Network_as_a_service
- https://en.wikipedia.org/wiki/Payments_as_a_service
- https://en.wikipedia.org/wiki/Recovery_as_a_service
- https://en.wikipedia.org/wiki/Security_as_a_service
- https://en.wikipedia.org/wiki/Unified_communications_as_a_service
- https://en.wikipedia.org/wiki/As_a_service
- https://en.wikipedia.org/wiki/Category:As_a_service

Ref: Wikipedia

Cloud Storage Example: S3

- Amazon Simple Storage Service (S3)
- Unlimited storage
- Pay for what you use

Region: Asia Pacific (Singapore)

	Storage pricing
S3 Standard - General purpose storage for any type of data, typically used for frequently accessed data	
First 50 TB / Month	$0.025 per GB
Next 450 TB / Month	$0.024 per GB
Over 500 TB / Month	$0.023 per GB
S3 Intelligent - Tiering * - Automatic cost savings for data with unknown or changing access patterns	
Frequent Access Tier, First 50 TB / Month	$0.025 per GB
Frequent Access Tier, Next 450 TB / Month	$0.024 per GB
Frequent Access Tier, Over 500 TB / Month	$0.023 per GB
Infrequent Access Tier, All Storage / Month	$0.02 per GB
Archive Access Tier, All Storage / Month	$0.005 per GB
Deep Archive Access Tier, All Storage / Month	$0.002 per GB
Monitoring and Automation, All Storage / Month	$0.0025 per 1,000 objects
S3 Standard - Infrequent Access * - For long lived but infrequently accessed data that needs millisecond access	
All Storage / Month	$0.02 per GB
S3 One Zone - Infrequent Access * - For re-createable infrequently accessed data that needs millisecond access	

Cloud Compute Example: EC2

- Amazon Elastic Compute Cloud (EC2)
  - Virtual computing environments (“instances”)
  - Pre-configured templates for instances
  - Launch as many virtual servers as needed (“elastic”)
  - Xen and KVM hypervisor

Amazon Elastic Compute Cloud (EC2)

Ref: https://en.wikipedia.org/wiki/Amazon_Elastic_Compute_Cloud

Do You Use The Cloud?

Cloud for IoT

AWS IoT

ThingSpeak

Samsung ARTIK™ Cloud

Microsoft Azure IoT Platform

IBM Watson IoT™

Summary

- Cloud Computing: Pros and Cons
- Hypervisors: Type 1 (bare metal) and Type 2 (hosted on another OS)
- Cloud Deployment Models: Public, Private, Hybrid etc.
- Cloud Service Models: IaaS, PaaS and SaaS
- Fog Computing: in between edge and cloud computing

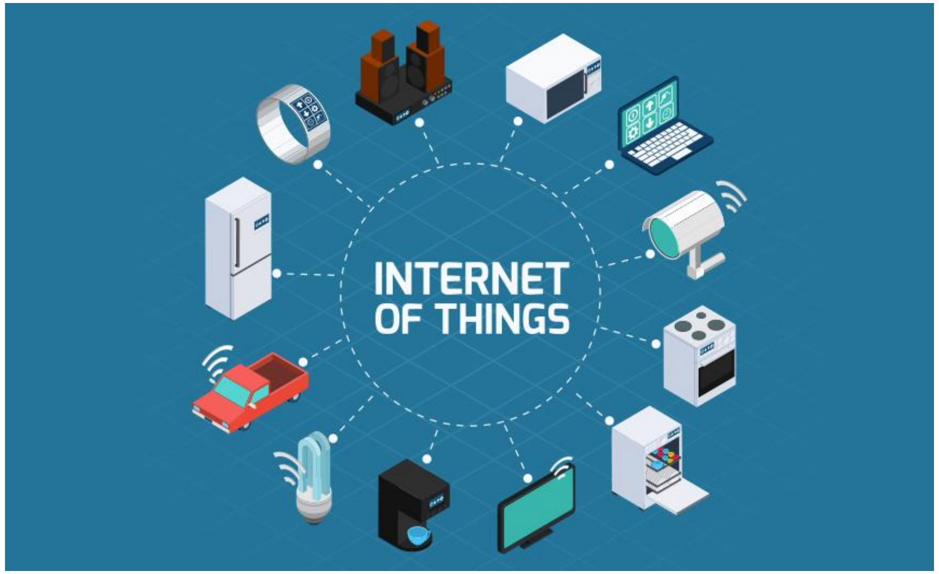
Internet of Things: Protocols and Networks (CSC2106)

AIoT: Artificial Intelligence on IoT

AIoT: Fusing AI with IoT – Relies on Edge Computing

SINGAPORE INSTITUTE OF TECHNOLOGY

Traditional IoTs

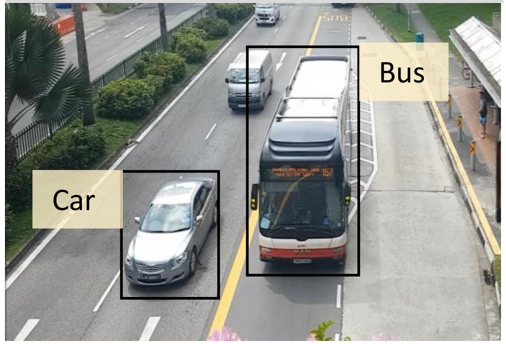


Source: https://www.bankinfosecurity.com/gao-assesses-iot-cybersecurity-other-risks-a-9926

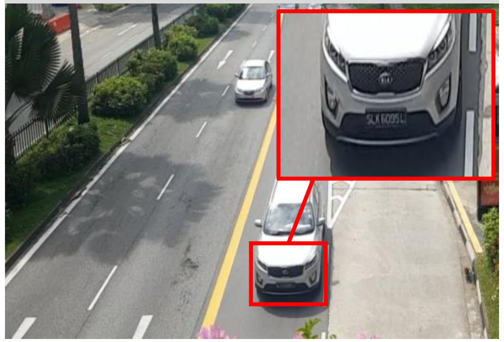
AIoTs

Crowd Counting

Object Localization & Classification

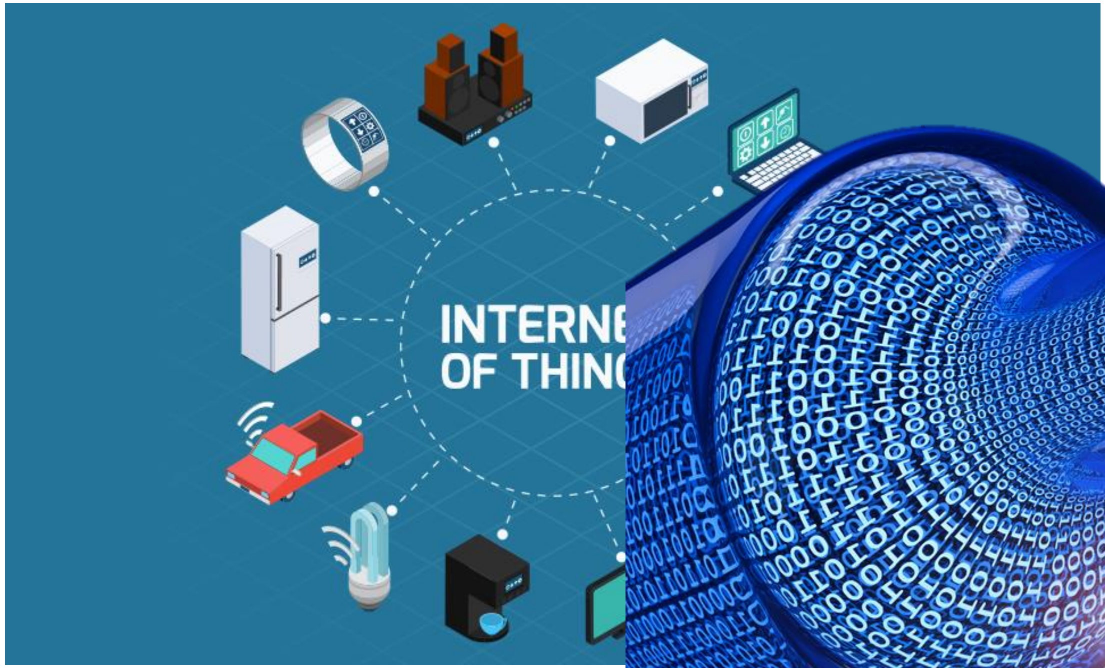


License Pate Recognition

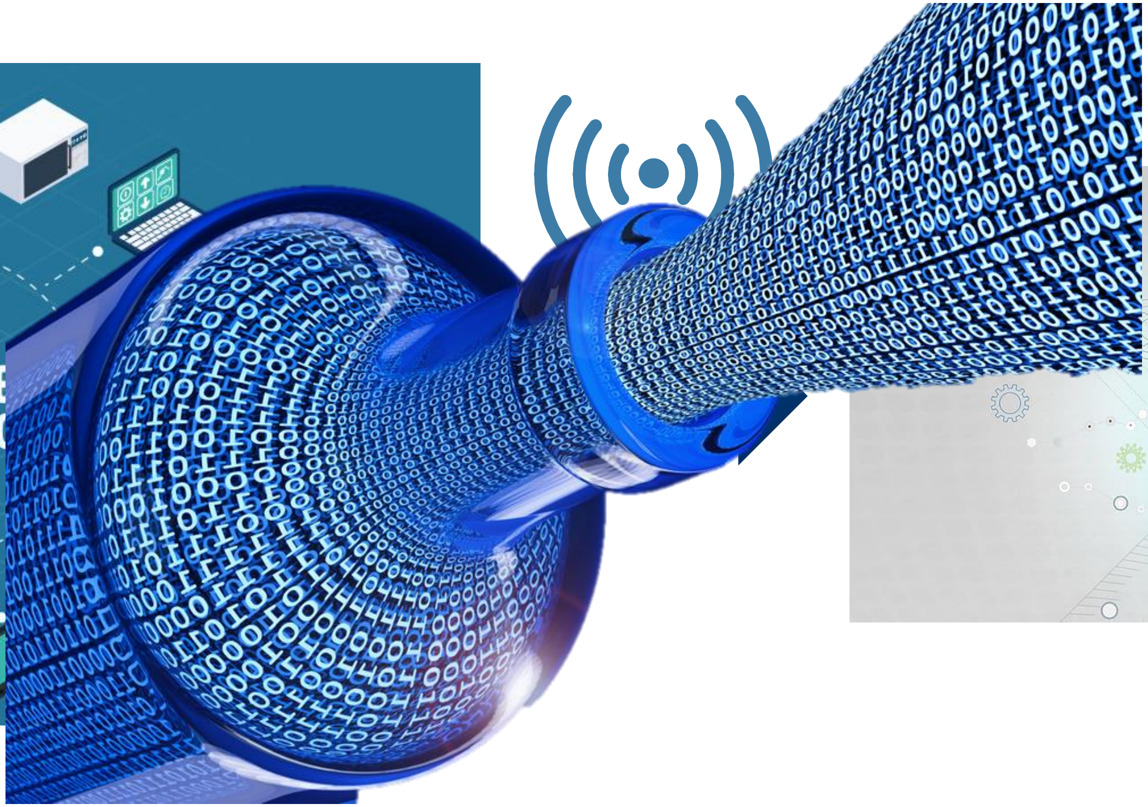


The Need for Edge Computing

- IDC estimates that there will be **41.6 billion connected IoT** devices generating a whopping **79.4 zettabytes of data** in 2025[1].



Source: https://www.bankinfosecurity.com/gao-assesses-iot-cybersecurity-other-risks-a-9926

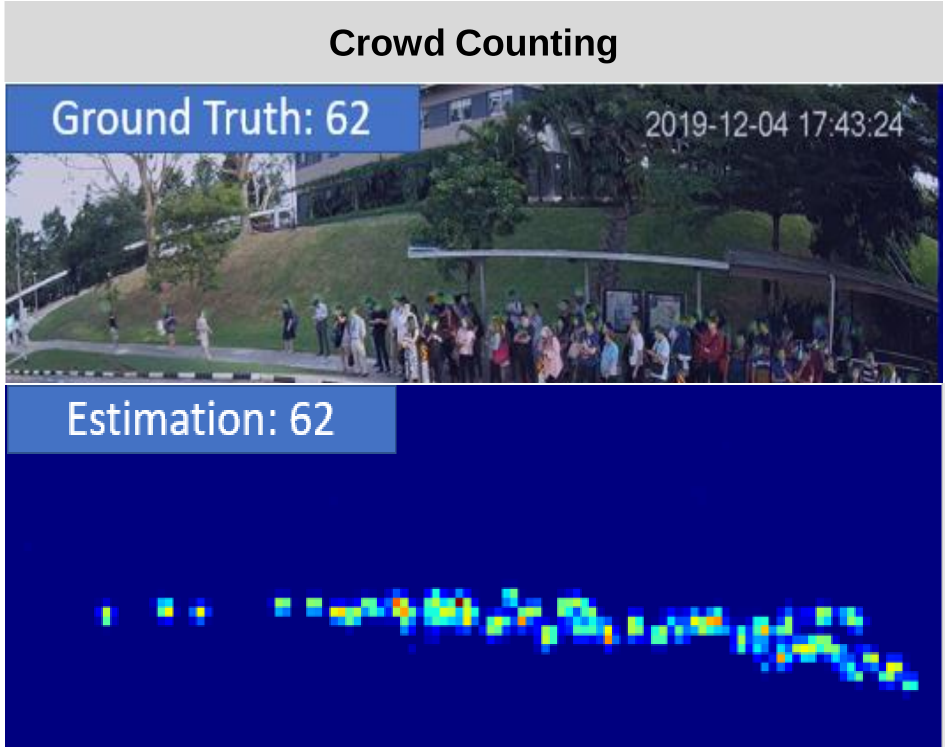


Source: https://www.hpcwire.com/

[1] Worldwide Global DataSphere IoT device and data forecast (2019-2023), IDC

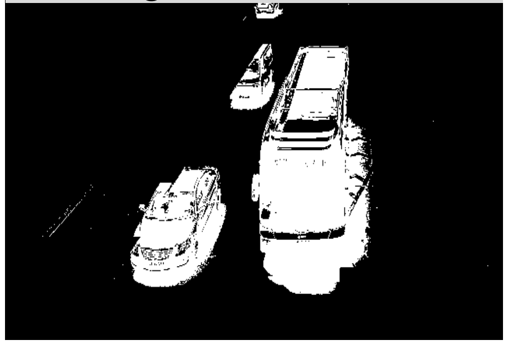
Advantages of Edge Computing

- Reduces Network Bandwidth Requirement
- Faster response time
- Significantly reduced data storage requirements
- …

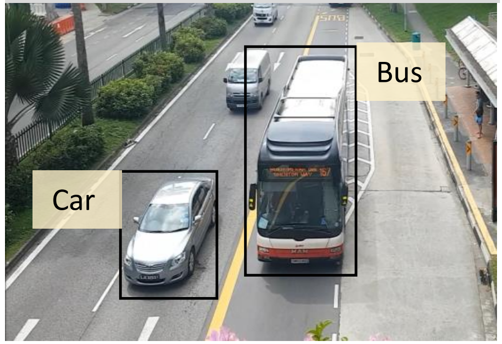


Example Application: Illegal Parking Detection

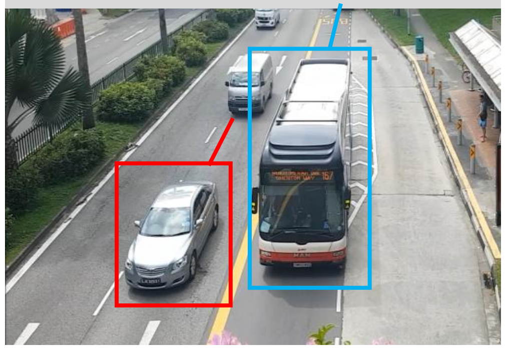
Pre-processing Background Modeling & Foreground Detection



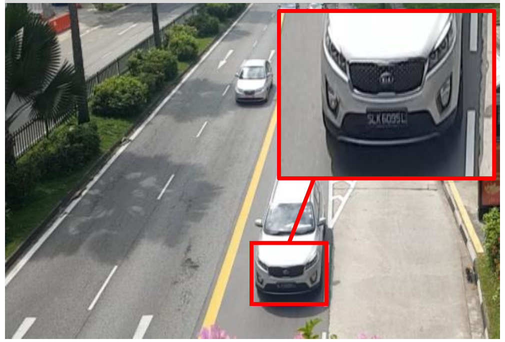
Classification



Vehicle Localization & Tracking

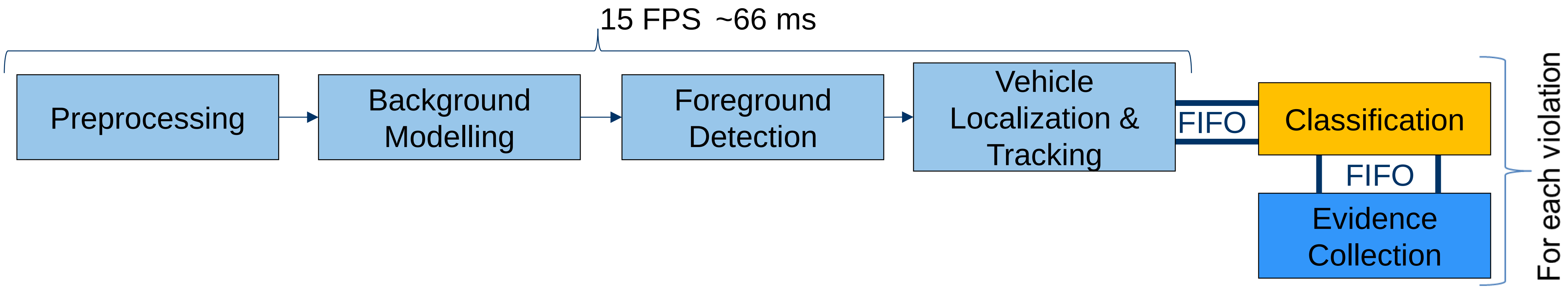


Evidence Collection



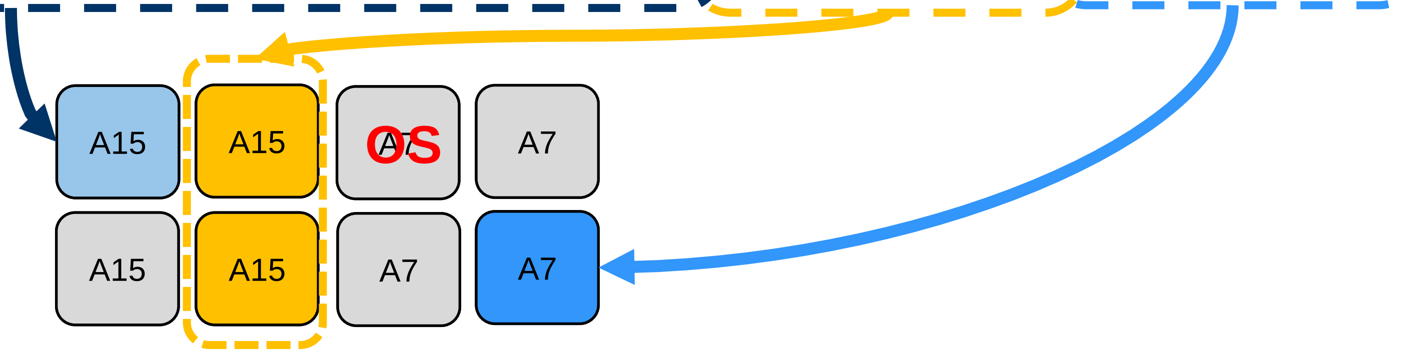
Performance Analysis and Mapping – Illegal Parking

SINGAPORE INSTITUTE OF TECHNOLOGY



Preprocessing	Background Modelling & Foreground detection	Vehicle Localization & Tracking	Classification	Evidence Collection
3.3 ms on one A15 30 ms on one A7	6 ms on one A15 50 ms on one A7	12 ms on one A15 80 ms on one A7	500 ms on one A15 350 ms on two A15	20 ms on one A15 100 ms on one A7

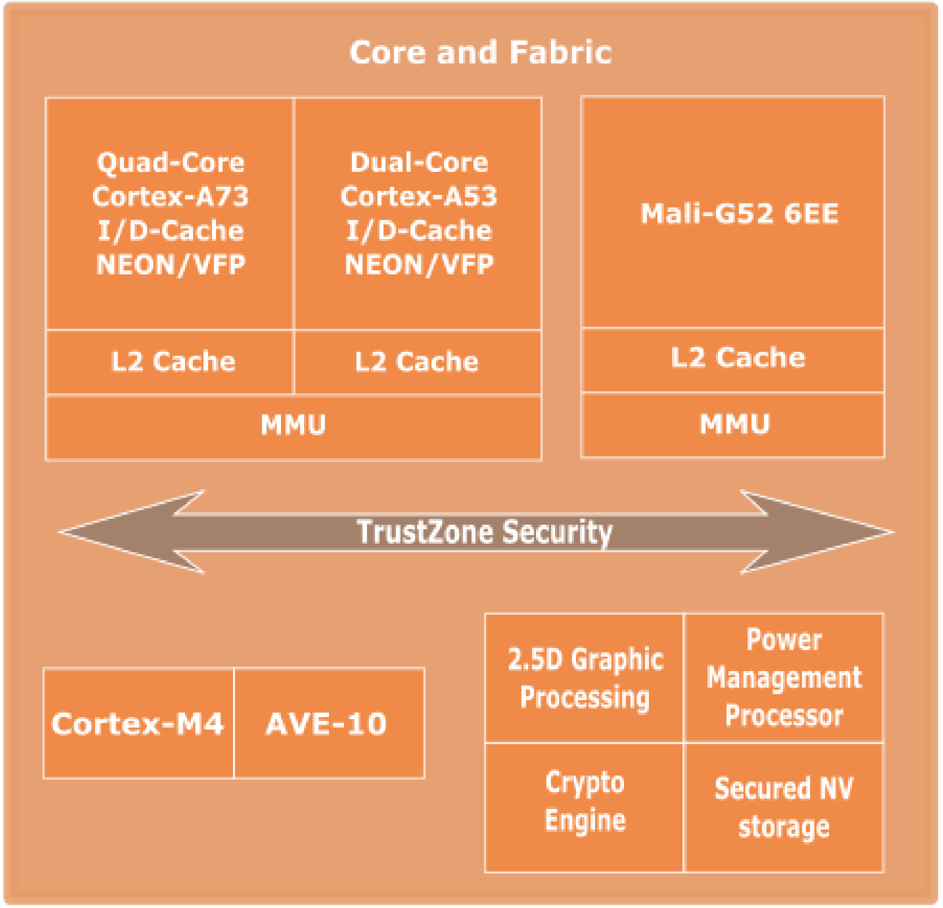
Given a power budget, a possible mapping could be:



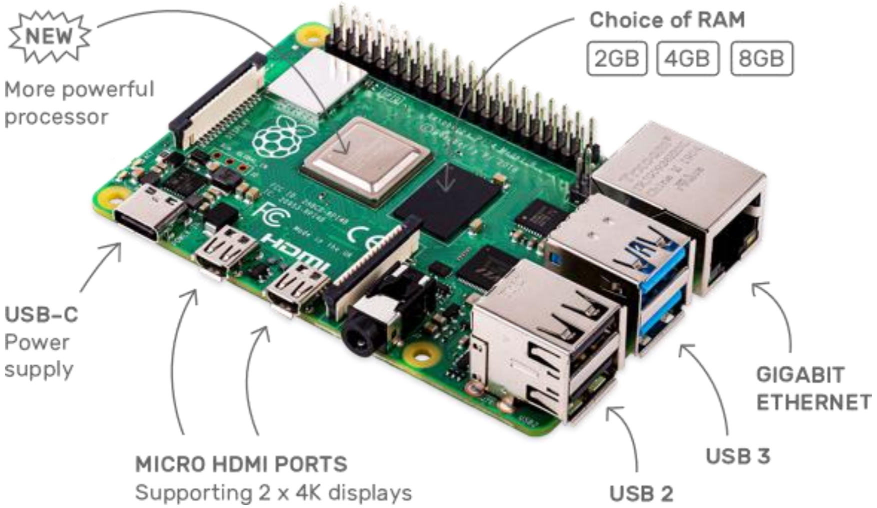
Edge Devices – CPU centric

Raspberry Pi 4

S922X



6-core CPU from Odroid N2+



Specifications

- Broadcom BCM2711, Quad core Cortex-A72 (ARM v8) 64-bit SoC @ 1.5GHz
- 2GB, 4GB or 8GB LPDDR4-3200 SDRAM (depending on model)
- 2.4 GHz and 5.0 GHz IEEE 802.11ac wireless, Bluetooth 5.0, BLE
- Gigabit Ethernet
- 2 USB 3.0 ports; 2 USB 2.0 ports.
- Raspberry Pi standard 40 pin GPIO header (fully backwards compatible with previous boards)

Edge Devices – GPU centric

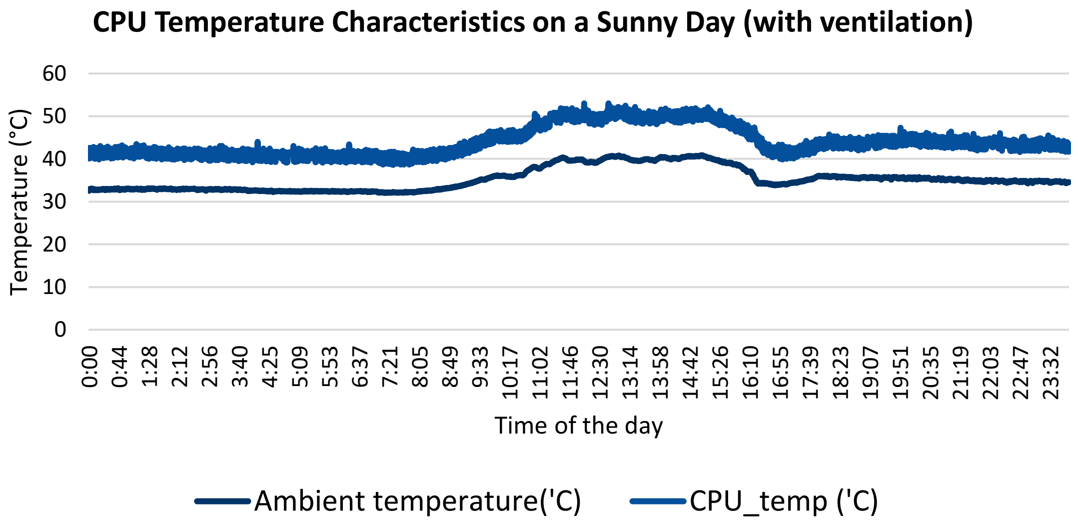
Compare NVIDIA Jetson Module Specifications

	Jetson Nano	Jetson TX2 Series				Jetson Xavier NX	Jetson AGX Xavier
		TX2 NX	TX2 4GB	TX2	TX2i		
AI Performance	472 GFLOPS	1.33 TFLOPS			1.26 TFLOPS	21 TOPS	32 TOPS
GPU	128-core NVIDIA Maxwell™ GPU	256-core NVIDIA Pascal™ GPU				384-core NVIDIA Volta™ GPU with 48 Tensor Cores	512-core NVIDIA Volta™ GPU with 64 Tensor Cores
CPU	Quad-core ARM® Cortex®-A57 MPCore processor	Dual-core Denver 2 64-bit CPU and quad-core Arm® Cortex®-A57 MPCore processor				6-core NVIDIA Carmel ARM®v8.2 64-bit CPU 6MB L2 + 4MB L3	8-core NVIDIA Carmel Arm®v8.2 64-bit CPU 8MB L2 + 4MB L3
Memory	4 GB 64-bit LPDDR4 25.6GB/s	4 GB 128-bit LPDDR4 51.2GB/s		8 GB 128-bit LPDDR4 59.7GB/s	8 GB 128-bit LPDDR4 (ECC Support) 51.2GB/s	8 GB 128-bit LPDDR4x 51.2GB/s	32 GB 256-bit LPDDR4x 136.5GB/s
Storage	16 GB eMMC 5.1	16 GB eMMC 5.1		32 GB eMMC 5.1	32 GB eMMC 5.1	16 GB eMMC 5.1	32GB eMMC 5.1
Power	5W | 10W	7.5W | 15W			10W | 20W	10W | 15W	10W | 15W | 30W

Real life Deployment Challenges

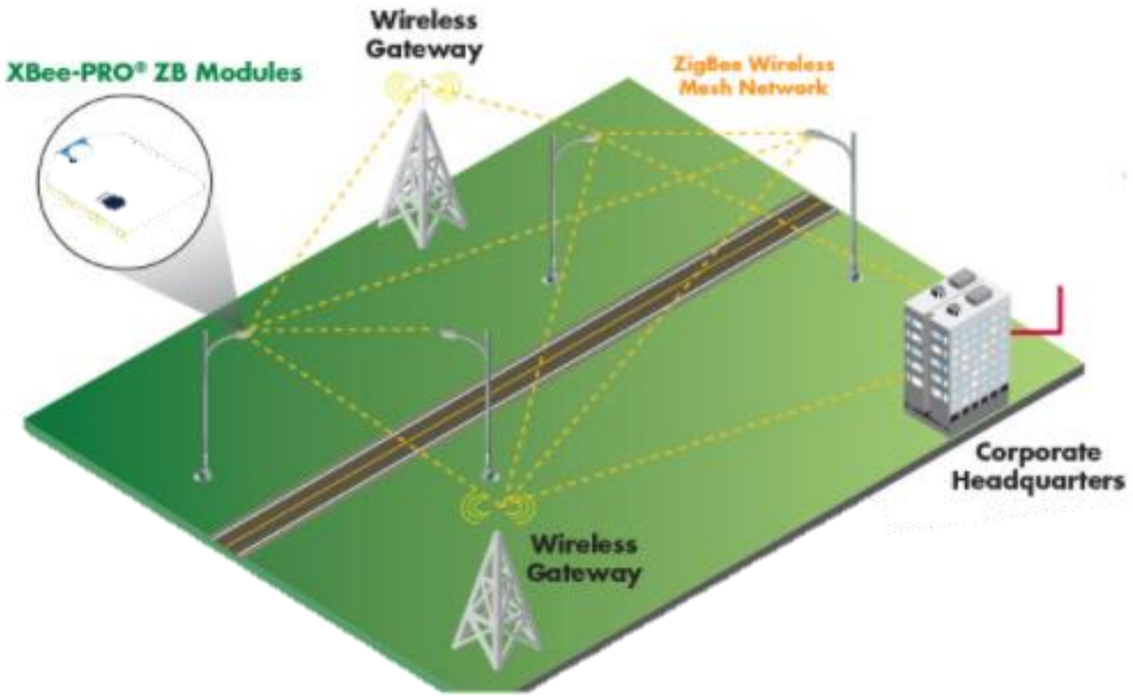
- Waterproofing requirements
- Thermal Challenges
- Waterproofing and thermal are conflicting requirements
- Temperature inside the box can rise over 65°C during daytime, without any ventilation
- And CPU Temperature is even higher than that. Can easily rise to 85°C during daytime

Traffic Junction



From Edge to Cloud: Networking Options

- Local edges can be connected to a gateway device through a localized interfacing protocol such as ZigBee, Z-wave and Wi-Fi.
- Connectivity to Cloud could be done using LTE, upcoming 5G or the myriad of other long range protocols we studied during this course, depending on the need for bandwidth, amount of data to be sent, cost of networking, etc.



EDGE AI ECOSYSTEM

DNN network architecture
- Classification (VGG, Mobilenet, Squeezenet)
- Object detection (faster-RCNN, YOLO, SSD)

DNN framework (training,inference)
Caffe, Caffe2, Pytorch (Facebook), Tensorflow (Keras)(Google), MXNet, Darknet

Inference-only DNN framework
TensorflowLite, NCNN, Pytorch Mobile
(Weight Compression, Filter Pruning, Quantisation)

CV and ML Hardware Libraries
- ARM NN and ARM Compute Library

CPU
- ARM big.LITTLE

GPU
- Nvidia Jetson
- ARM Mali

SoC Families
Hisilicon Kirin, Samsung Exynos, Qualcomm Snapdragon, Nvidia Jetson, Apple Bionic

FPGA/Accelerator
Google edge TPU (coral), Intel Nirvana NNP, Apple Neural Engine, Huawei NPU (in Kirin SoC)

Hardware Programming Languages

Multi-core CPU
- OpenMP, NEON (extension)

GPU
- CUDA, OpenCL, cuDNN

Manufacturing Revolution – From Industry 1.0 to Industry 4.0

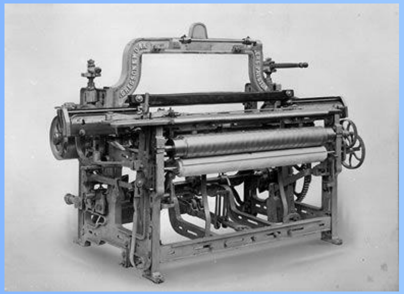
SINGAPORE INSTITUTE OF TECHNOLOGY

Industry 1.0

FIRST

Industrial Revolution

Key Change:
Introduction of Mechanical Production Equipment driven by Water and Stream Power



18th Century Mechanical Loom

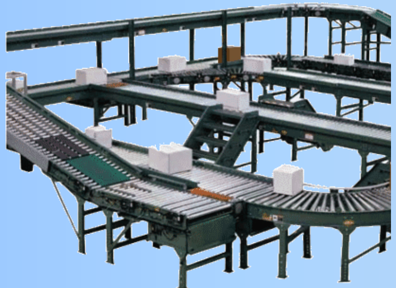
End of 18th Century

Industry 2.0

SECOND

Industrial Revolution

Key Change:
Introduction of mass Manufacturing Production lines powered by Electric Energy



Vintage Electric Conveyor Belt

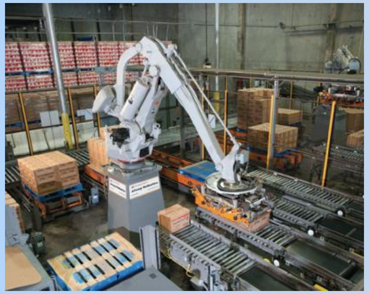
End of 19th Century

Industry 3.0

THIRD

Industrial Revolution

Key Change:
Introduction of Electronics, PLC Devices, Robots and IT to automate Production



PLC Driven Robots

Q4 of 20th Century

Industry 4.0

FOURTH

Industrial Revolution

Key Change:
Introduction of IoT and Cyber-Physical Systems driven by Augmented Reality & Real Time Intelligence

Augmented Reality Driven CPS

Start of 21st Century

Level Of Complexity

Source: Entrigna

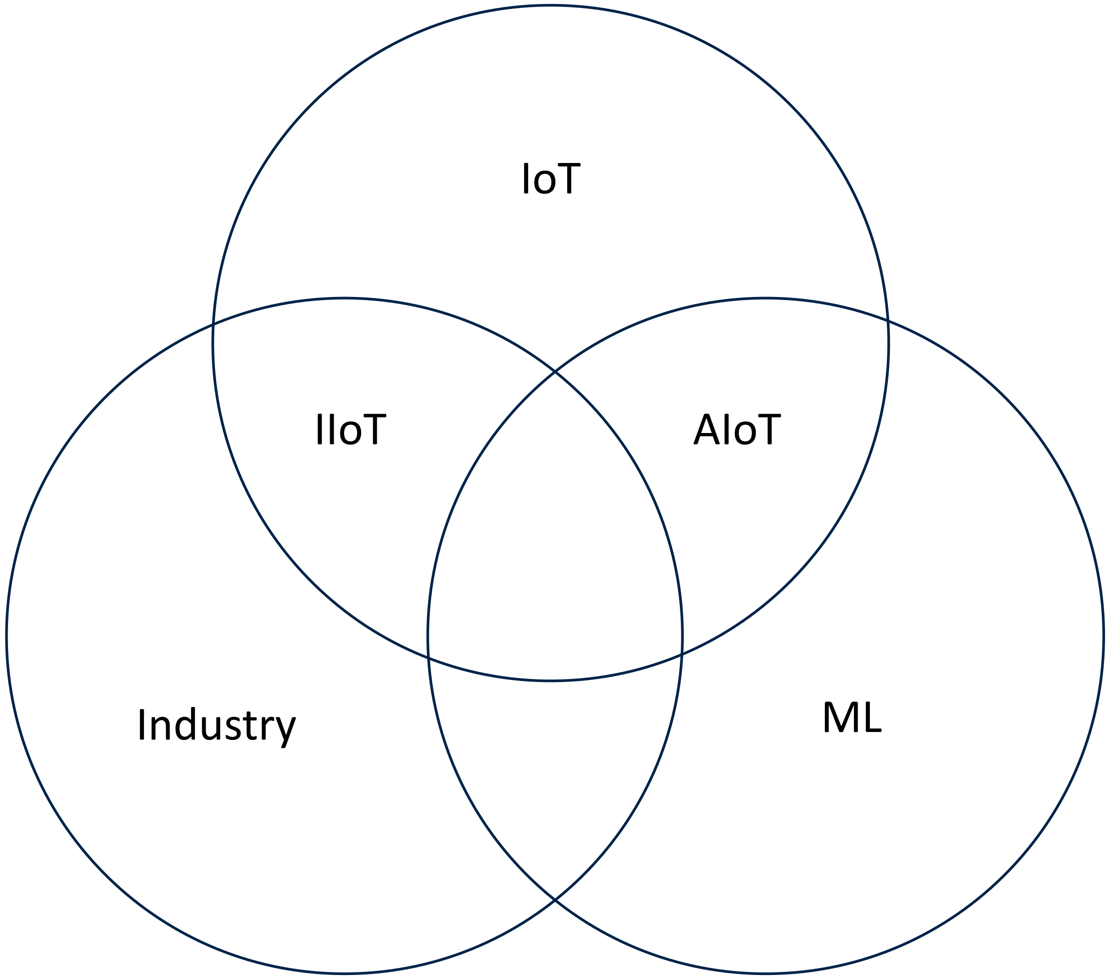
IoT vs. IIoT

APPLICATION DOMAINS

ACCURACY AND PRECISION ARE TOP PRIORITIES FOR IIOT

CYBER SECURITY IS CRITICAL

RELIABILITY AND ROBUSTNESS


IoT
IIoT
AIoT
Industry
ML

Smart Manufacturing Examples: Airbus – Factory of the Future

SINGAPORE INSTITUTE OF TECHNOLOGY

- MiRA (Mixed Reality Application) tablet
  - Cross between a sensor pack and a tablet

- Internet Connected Smart Tools
  - Auto-adjust to different actions
  - Log information
  - Reduces assembly time

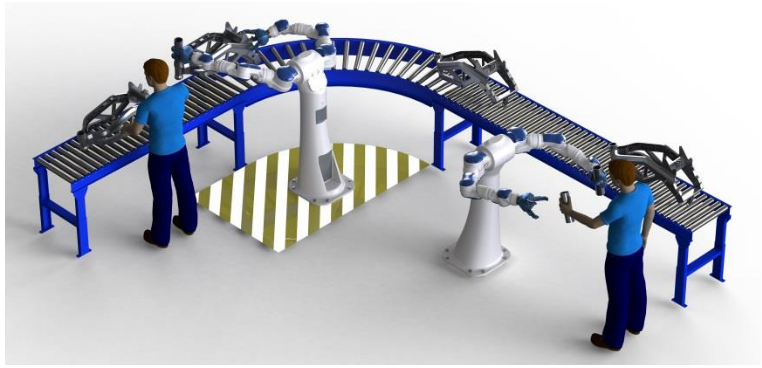
- Augmented Reality driven instructional & educational tutorials

Photos courtesy of [Airbus Factory of the Future]

Smart Manufacturing Examples: Continental AG's SMART Factory

- Active RFID tags and Geo-location are used to move the tire components throughout the factory

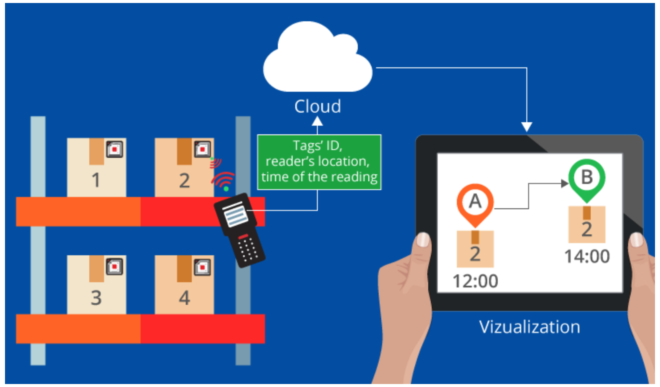
- Collaborative robots
  - Robots are "shown" how to do a task once and then they can repeat that action
  - Reduces risks of injuries and reduces the need for additional assisting employees

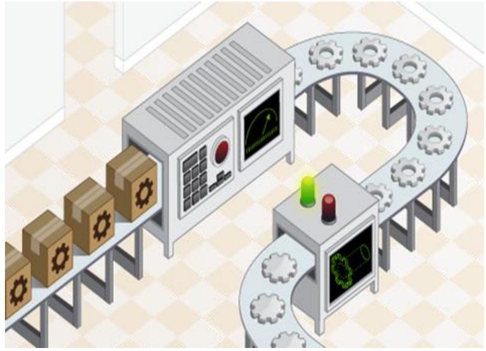


Source: Entrigna

Going forward

- Streamlined Factories
  - Asset tagging, locating, supply chain management
  - Easier to do Just-in-time asset management
- SMART Inventory management
  - Sensors on containers can determine when a product is running low
  - Automatic alerts to proactively re-order the parts or orders can be automatically placed with suppliers
- SMART Quality control
  - RFIDs attached to products can be used to tag defective products
  - Automatic alarms if failure rate crosses a certain threshold for early course correction





Source: Entrigna

Summary

- AIoT: Artificial Intelligence Tasks on IoT Devices
  - Only the Inference task typically done on the IoT
  - Training is still performed offline on desktop class GPUs
- Several Edge Computing Platforms
  - CPU centric: e.g., Odroid boards, Raspberry Pi, etc.
  - GPU centric: e.g., several Nvidia Jetson platforms
- Drastically reduces the communication load.
- Will be increasingly important as we connect more and more devices that will end up generating even more data.
- Industrial IoT: IoTs used in smart manufacturing, asset tagging, locating, supply chain, etc..
- IIoTs need to satisfy stricter constraints in terms of accuracy, reliability, cyber security, etc..