Cloud, Fog Computing for IoT and Data Analytics

Instructor: Deepak Gangadharan

Outline

- Cloud Computing
- Fog Computing
- Data Analytics for IoT

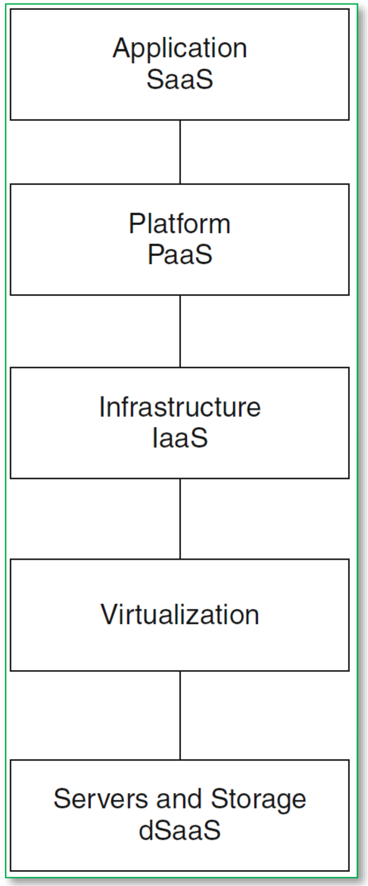
Cloud Computing – An Introduction

- **Definition:** A style of computing in which dynamically scalable and often virtualized resources provided as services over the Internet
- Variety of devices such as PCs, laptops, smartphones, etc. used to access programs, storage and application development platforms via services offered over the internet
- Advantages: Cost savings, High availability and Scalability

http://www.techeconomy.it/2013/07/24/la-privacy-dei-dati-nellera-del-cloud-computing/

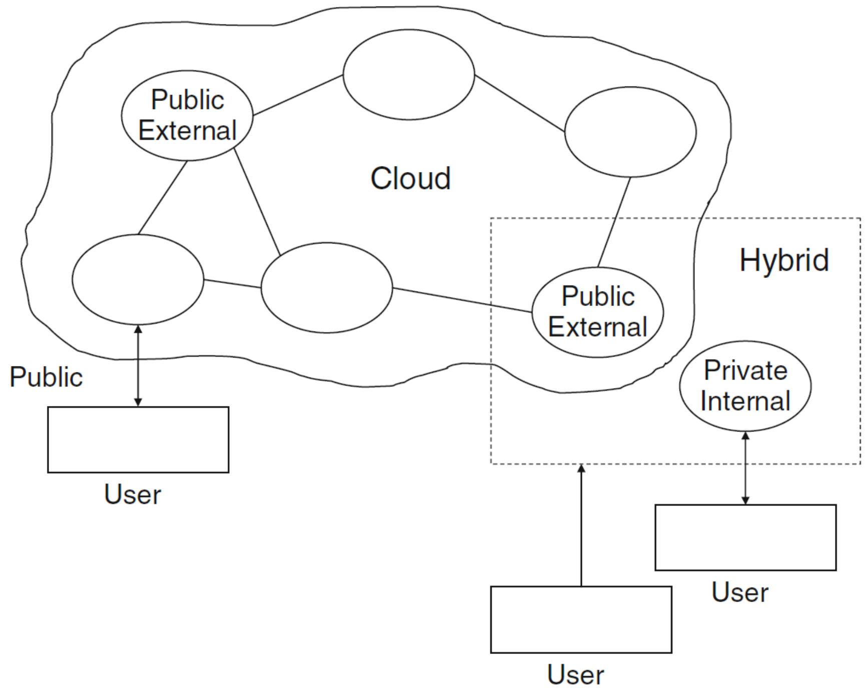
Layered Architecture

- data-Storage-as-a-Service (dSaaS)
  - Provides storage that a consumer uses including bandwidth for storage
- Infrastructure-as-a-Service (IaaS)
  - Computing resources provided as a service – virtualized compute nodes with guaranteed processing and reserved bandwidth
- Platform-as-a-Service (PaaS)
  - IaaS + OS + required services for an application
  - Example PaaS providers: Google App Engine, Microsoft Azure
- Software-as-a-Service (SaaS)
  - Allows users to run applications remotely
  - Example: Google Apps, Microsoft Office 365



Types of Cloud Computing

- Public cloud (external cloud)
  - Computing resources provisioned from an off-site third party provider
  - Applications from different users likely to be mixed together
- Private cloud (internal cloud)
  - Built for exclusive use of one client
  - Full control over data, security and quality of service
  - Built and managed by a company's IT organization or a cloud provider
- Hybrid cloud
  - Combines multiple public and private cloud models
  - Complexity in determining how to distribute applications



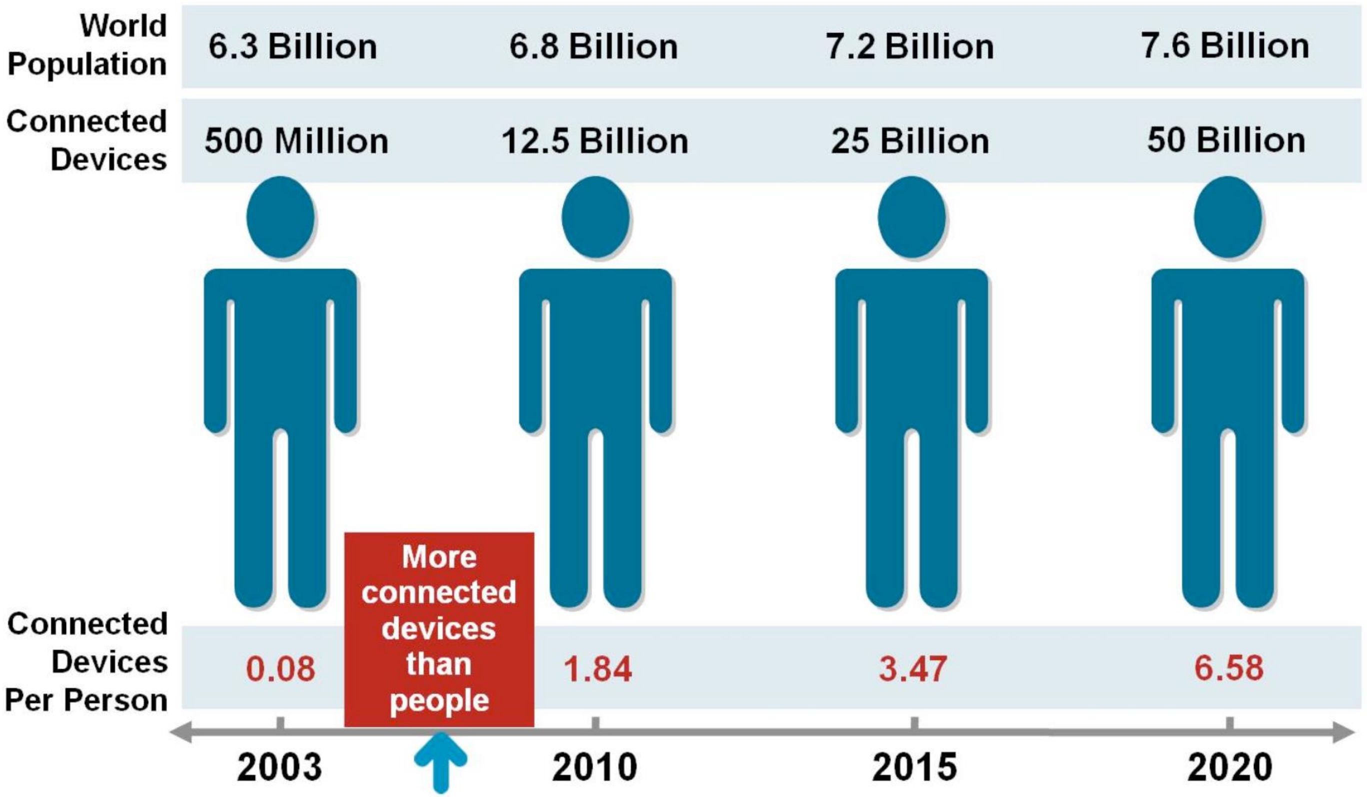
Features of Cloud Computing

- Scalability and on-demand services
    - Provides resources and services on-demand
    - Resources scalable over several data centers
- User-centric interface
    - Can be accessed by Web services and Internet browsers
- Guaranteed Quality of Service (QoS)
    - Guarantees QoS for users in terms of hardware/CPU performance, bandwidth and memory capacity
- Pricing
    - Users pay for services and capacity based on the need

Challenges in Cloud Computing

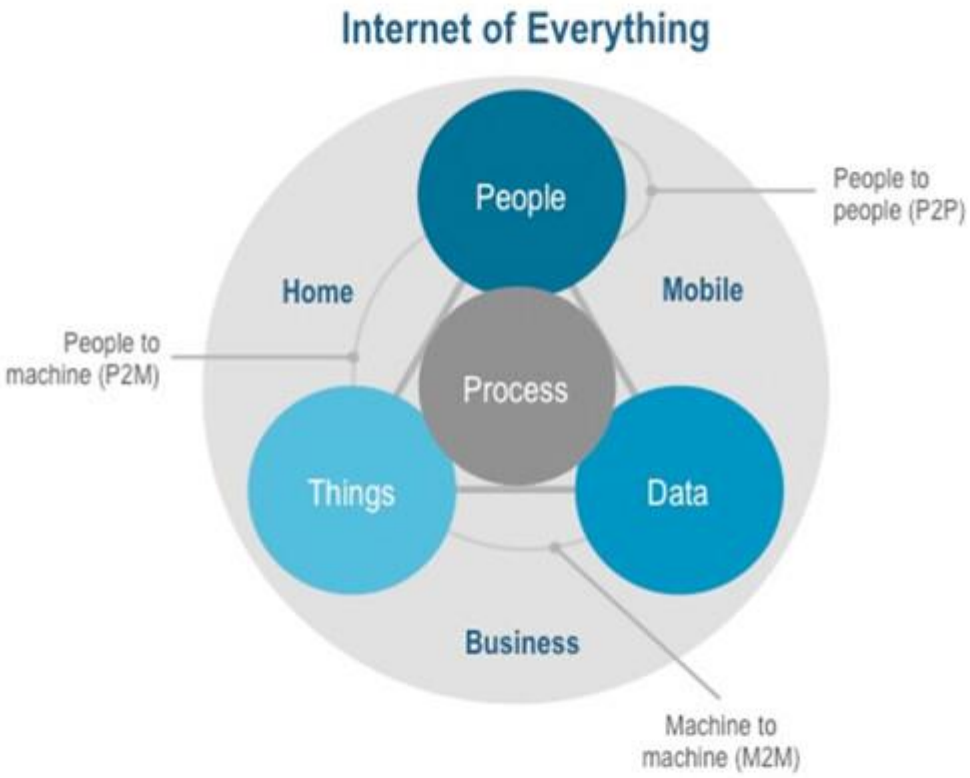
- Performance
  - Users far away from cloud providers may experience high latency
- Security and Privacy
  - Concerns over vulnerability of attacks, when information and critical IT resources are outside firewall
  - New attack surfaces possible between VMs
- Control
- Bandwidth costs
- Reliability

Emergence of IoT



Source: The Internet of Things by Dave Evans, CISCO IBSG 2011

Emergence of Internet of Everything



Source: https://live.staticflickr.com/8490/8178840985_22dc5fa9dd.jpg

Challenges of Cloud Computing

Large data volume

High Latency

Resiliency impractical

High transport cost

Motivation

- There will be 35 billion IoT devices worldwide by 2021
- Data generated by IoT devices will be 79.4 zettabytes by 2025 (Requires 1000 data centers!)
- Cloud Computing Characteristics
  - High latency due to significant distance between end devices and cloud
  - Centralized model of computing
  - No location awareness
  - Poor support for mobility
  - Vulnerable to security attacks
- Issues are Volume, Latency and Bandwidth

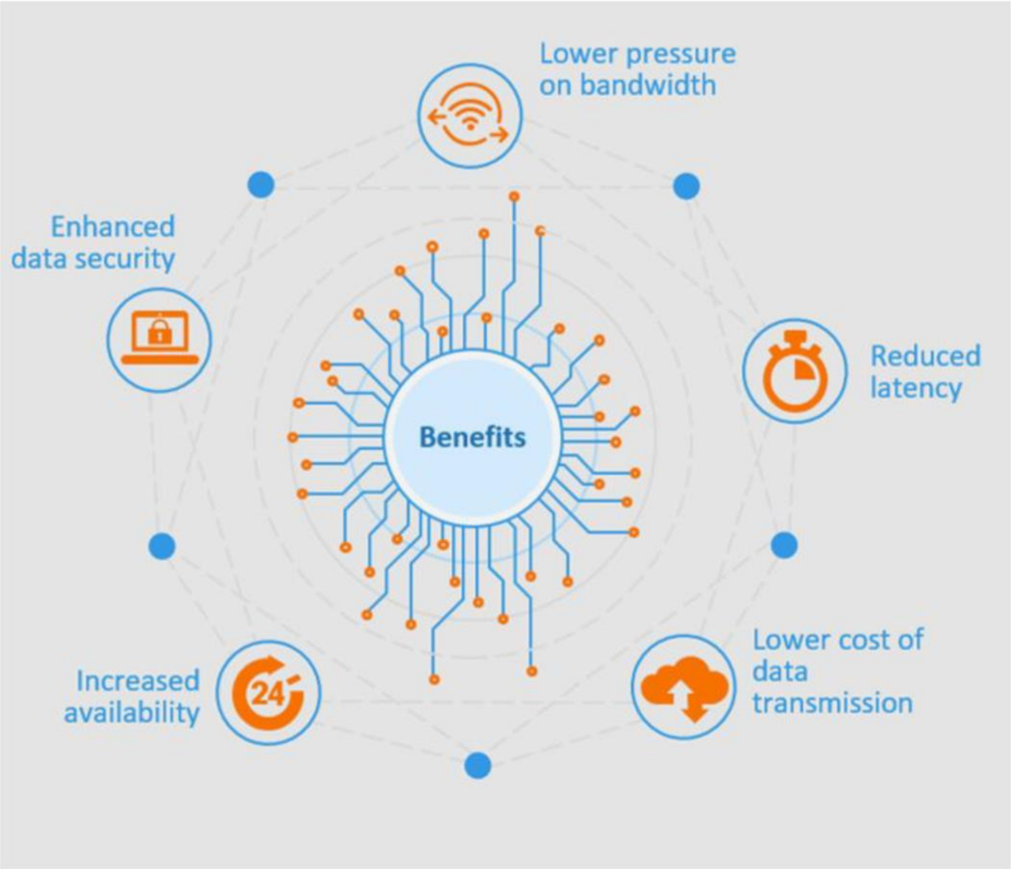
What is Edge Computing

- A distributed information technology (IT) architecture in which client data is processed at the periphery of the network, as close to the the originating source as possible.
- This new infrastructure involves sensors to collect data and edge servers to securely process data in real-time on site, while also connecting other devices, like laptops and smartphones, to the network.

Need of Edge Computing

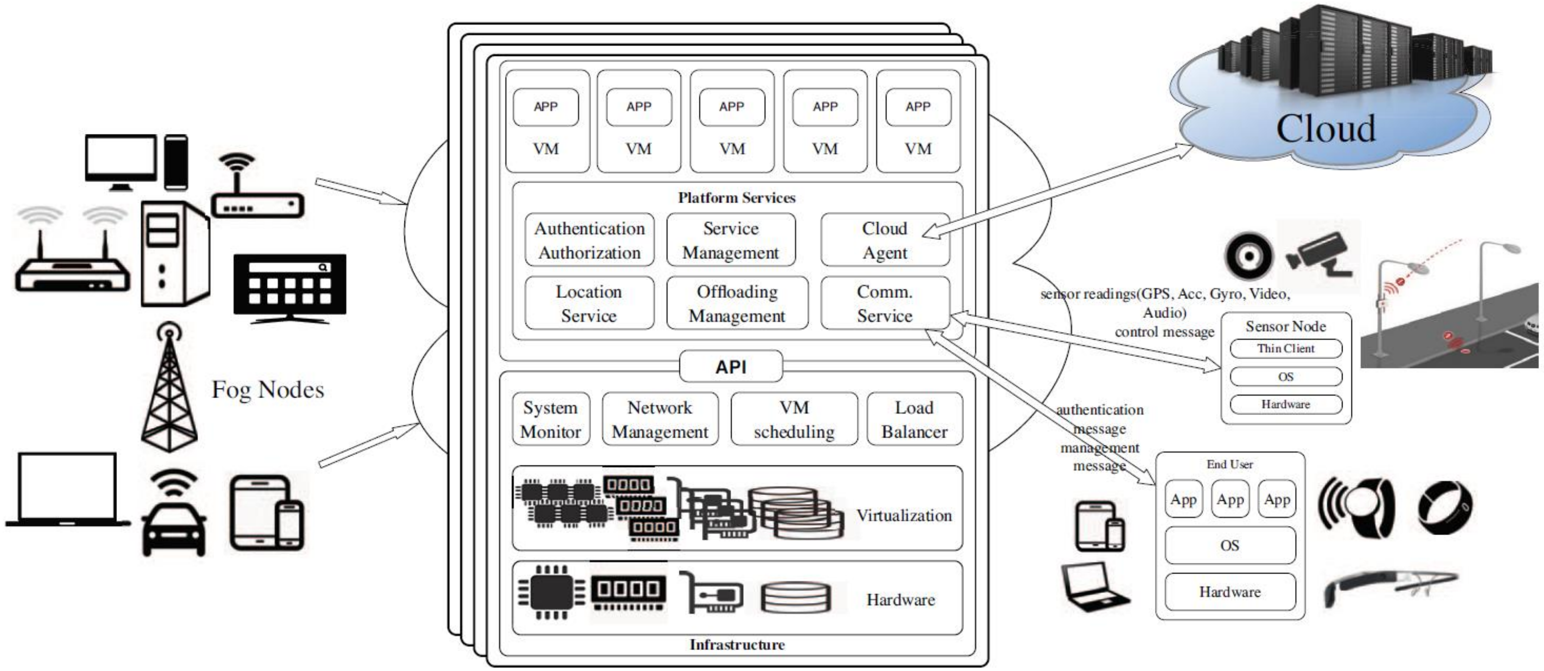
- Edge computing addresses vital infrastructure challenges such as
  - Decreasing the latency in data transmission/processing
  - Reducing the amount of data transmitted and stored in the cloud.
- Resolves the issues of Cloud computing via three Edge computing models - Fog computing, Mobile Edge computing and Cloudlets by limiting bandwidth, excess latency and network congestion

Characteristics and benefits of Edge Computing



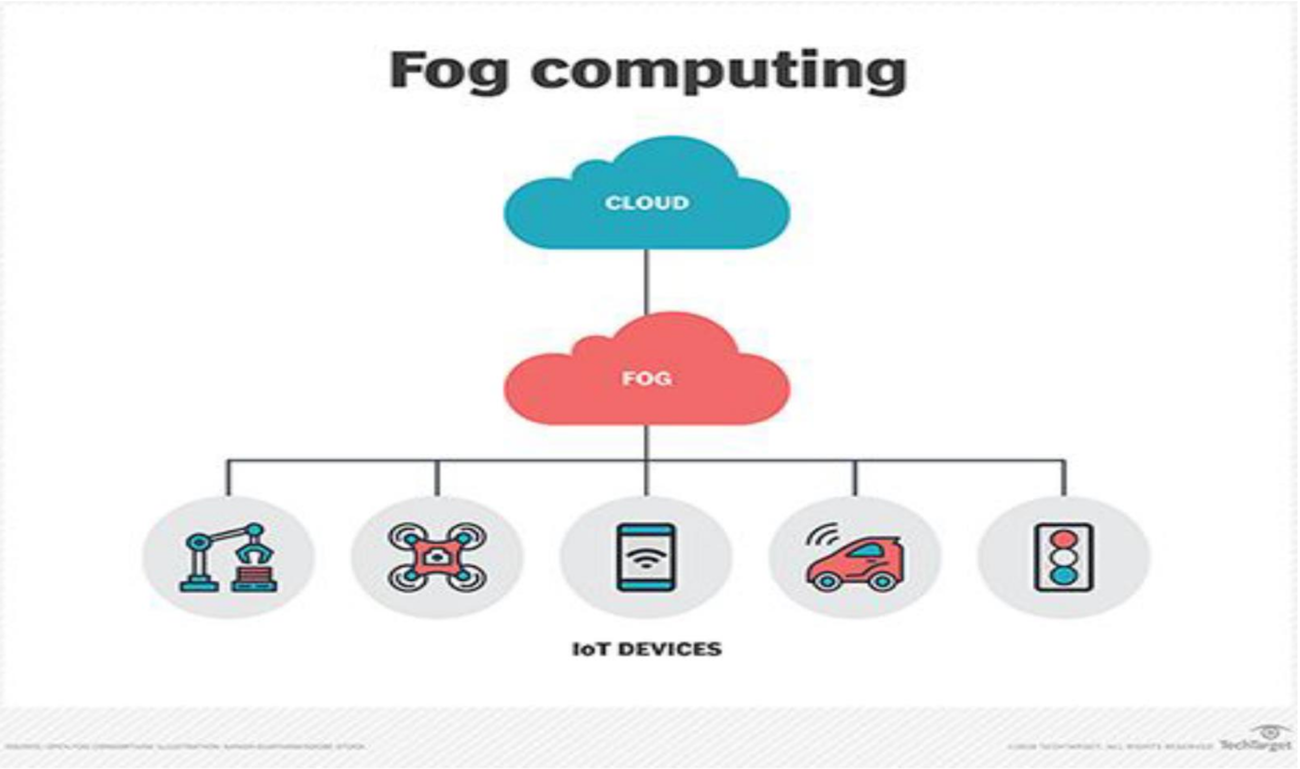
Source: https://www.everestgrp.com/2019-05-edge-computing-characteristics-and-benefits-market-insights-50131.html/

Components for Fog Computing Platform



Fog Computing

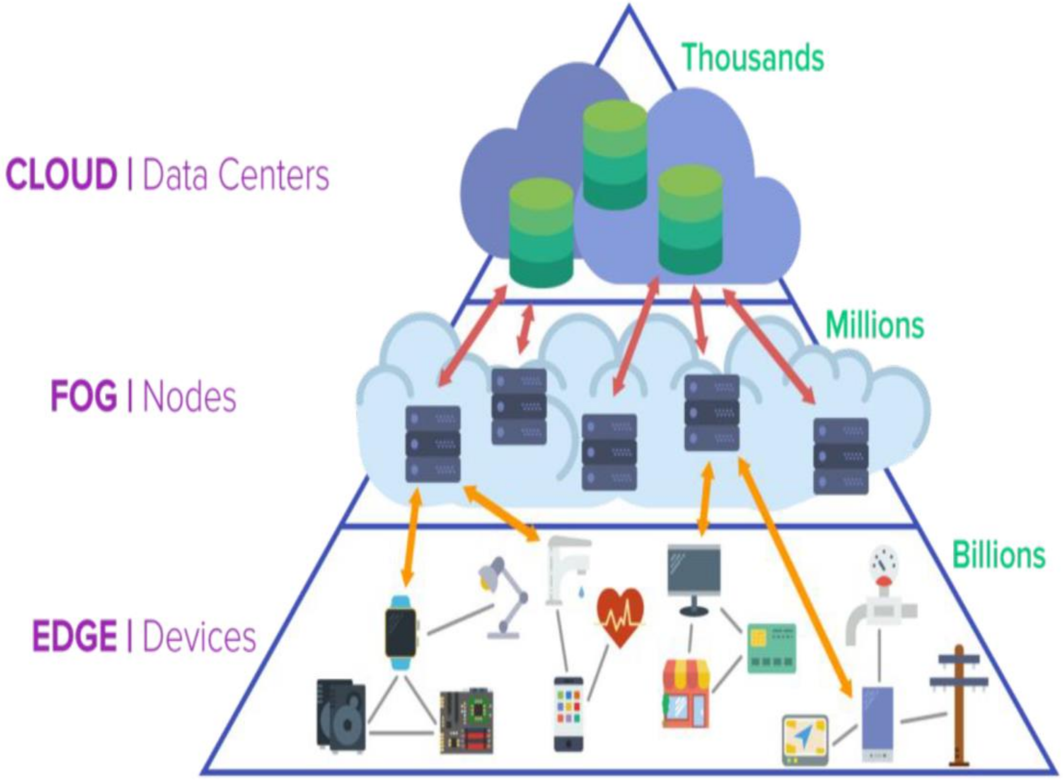
- The fog metaphor comes from the term for a cloud close to the ground, just as fog concentrates on the edge of the network.
- Fog computing is not meant to replace cloud computing, but to enhance its capabilities and improve performance.
- Fog computing introduces an intermediate layer called fog that is designed to process the communication data between the cloud and end users.



Source : https://internetofthingsagenda.techtarget.com/definition/fog-computing-fogging

Difference between Fog and Edge Computing

- The fog metaphor comes from the term for a cloud close to the ground, just as fog concentrates on the edge of the network.
- Fog computing is not meant to replace cloud computing, but to enhance its capabilities and improve performance.
- Fog computing introduces an intermediate layer called fog that is designed to process the communication data between the cloud and end users.



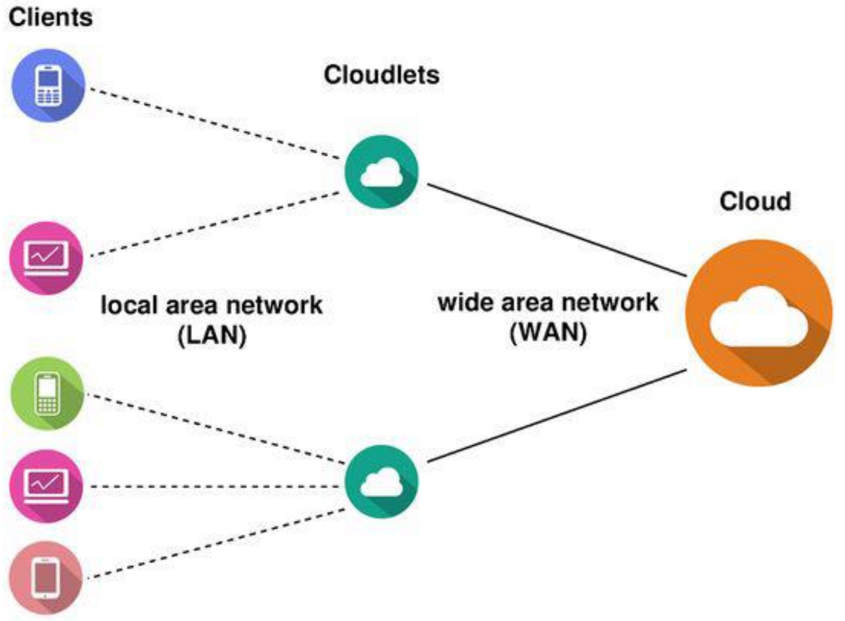
Source : https://erpinnews.com/fog-computing-vs-edge-computing

Benefits of Fog Computing

- Minimal amount of data sent to the cloud.
- Saves bandwidth as now all the unwanted data is not sent to the cloud.
- Reduces data latency.
- Improves data security (when limited data goes to the cloud and it is local processing, easy to protect)

Cloudlets

- Several small-scale data centers (DCs)
- Located on the edge of internet
- Well connected with the cloud
- Wireless connection to mobile devices



Source: https://thecustomizewindows.com/2020/05/what-is-a-cloudlet-in-cloud-computing/

Data Analytics in IoT

Introduction

- Value of IoT lies in the interpretation and decision made of the vast amounts of data.
- Data Analytics provides us with the valuable patterns within the data.
- Deals with structured data (SQL storage), unstructured data (raw video data and semi-structured data (twitter feeds).
- Data may need to be interpreted and analyzed in real time or may be archived and retrieved for deep analytics in the cloud

Basic Data Analytics in IoT

- Preprocessing:
  Filter events of little interest, feature extraction, segmentation, data transformation, adding attributes
- Alerting:
  Inspect and alert if data exceeds boundary condition
- Windowing:
  A sliding window of events is created – Time based or Length based
  Example: No. of spikes in any parameter in the last 1 hour
- Joins:
  Combines multiple data streams into a new single stream
- Errors:
  Detecting missing data, garbled data, and data that is out of sequence

Basic Data Analytics in IoT

- Temporal events and patterns:
  Sequence of events constituting a pattern of interest
- Trends
  Useful for predictive maintenance.
  A rule to detect an event based on time-correlated series data

Analytics in the Cloud

- Rules Engines:
  Define an action and produce an outcome based on events
- Stream Processing:
  A graph-based data processing architecture where nodes represent operators on data input and events sent to other operators.
  Graph can be replicated and executed in parallel
- Complex Event Processing:
  Often used for pattern detection.
  Many events reduced and transformed into higher-level events → more abstract than raw sensor data

Machine Learning in IoT

- Each machine learning model has its particular strength and use case it serves
- Goal is to predict or infer what a data set indicates
- Types of Learning Systems – Supervised, Unsupervised and semi-supervised learning
- Three fundamental uses of machine learning – Classification, Regression and Anomaly Detection.

Classification

- Form of supervised learning where data picks a category
- Two variants of classification:
  - Binomial: Picking between one of two categories
  - Multi-class: More than two categories
- Linear Classifier: Hyperplane dividing the two categories is a straight line.
- Nonlinear relationships causes severe errors with a linear model.
- A non-linear model tends to overfit the test series

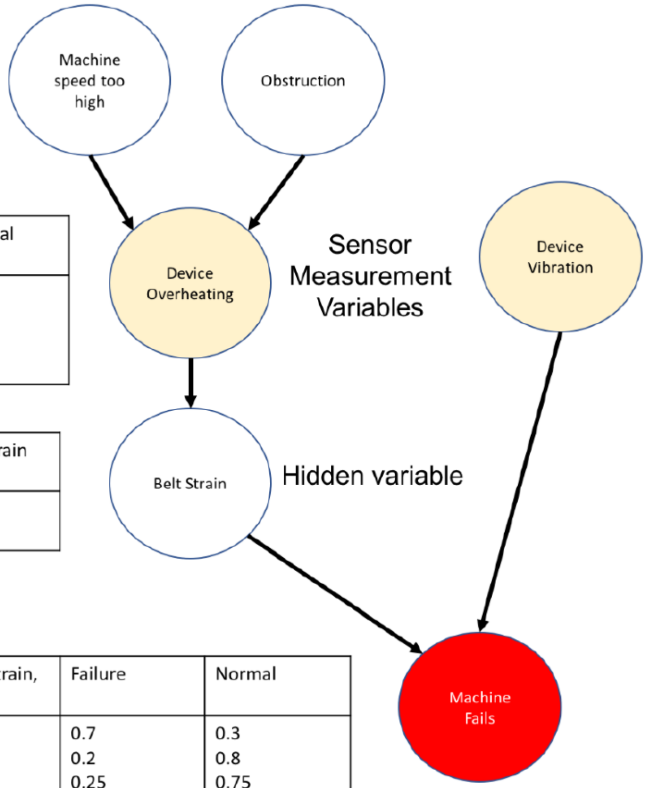
Bayesian Models

- Based on Baye's theorem
- Probability that an event will occur based on prior knowledge of the system – Example: Probability that a machine will fail based on temperature of the device
- Baye's Theorem expressed as
  $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)}$, where A and B are events of interest

Bayesian Network Model

P(Obstruction)	Obstructed	Normal
	0.1	0.9

P(Speed)	High	Normal
	0.01	0.99



P(Overheating| machine speed, obstruction)	Abnormal	Normal
High, obstructed	0.99	0.01
High, no obstruction	0.8	0.2
Normal, obstructed	0.7	0.3
Normal, no obstruction	0.05	0.95

P(Vibration Level Exceeds 10 Hz)	Abnormal	Normal
	0.02	0.98

P(Belt Strain | Temperature)	Strain	No Strain
Normal	0.05	0.95
Abnormal	0.7	0.3

P(Machine Failure | Belt Strain, Vibration Level)	Failure	Normal
Strain, Abnormal	0.7	0.3
Strain, Normal	0.2	0.8
No Strain, Abnormal	0.25	0.75
No Strain, Normal	0.05	095

Thank You