Cloud Computing Introduction and Foundations Concepts

Evolution towards Cloud Computing

Reference:

- [cam-san] Chapter 2

Cloud Computing - Definition

- **NIST, the US National Institute of Standards and Technology**: "*Cloud Computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort of service provider interactions*"

- **My Mom**: "The place where my phone saves my emails, contacts, pictures, so I don't lose them"

Evolution of technology

- Cloud computing is not an abrupt innovation, instead, **it is the result of a series of developments that have taken place over the past few decades**
- The concept traces back to 1960, when McCarthy (the same who coined the term ‘artificial intelligence’) wrote:
  - ‘*computation may someday be organized as public utility*’, like water, electricity, gas, etc...
- It took, however, more than 40 years for technology to develop and mature for cloud computing: *only the composition of some key technologies has enabled cloud computing*
- Several decades of research, especially in the domain of parallel and distributed computing, have paved the way

The origin – Mainframes

- Commercial usage of computing started around 1970s with mainframes
- Organizations acquired such hardware to automate basic data processing, e.g. payroll management
- Mainframes were large supercomputers installed in a room. They were so high priced that it *was hard for a company to have more than one*
- Multiple users were sharing the mainframe at the same time: the common approach was to implement a time-shared utilization
- Different users were accessing the mainframe from terminals

The origin – Mainframes

- Terminals were dumb terminals, i.e. were only capable of transferring input-output data to/from the mainframe (the server)
- This centralized processing environment was a bottleneck
- Users had always to wait for a long time in queue

PC Revolution

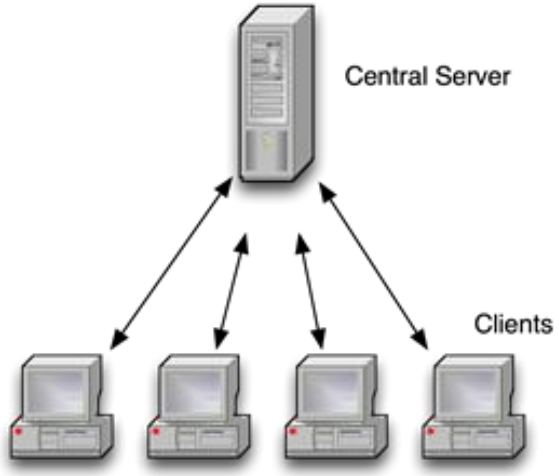
- At the end of the 70s less expensive computers came into the market
- Those Personal Computers (PCs) provided some processing and storage capabilities to run applications
- The need to maintain expensive mainframes systems soon became less important

Network of PCs

- Big organizations installed multiple PCs in different offices
- Each one, however, was functioning independently
- Network communication technologies were introduced to allow communication between PCs for data transfer
- LAN: Local Area Network (LAN) and Wide Area Network (WAN) were introduced at that time: *computer in the same office or placed at distance could communicate*
- Over time communication technologies improved, increasing the bandwidth

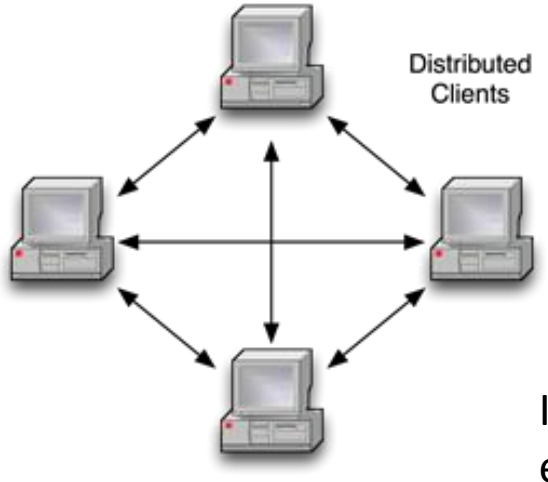
Distributed applications

- *Client/Server computing*: we have a server (offering some service/information) and a client



Client / Server

- *Peer to peer computing*: each and every computer can play the role of a client as well as a server



Peer to Peer

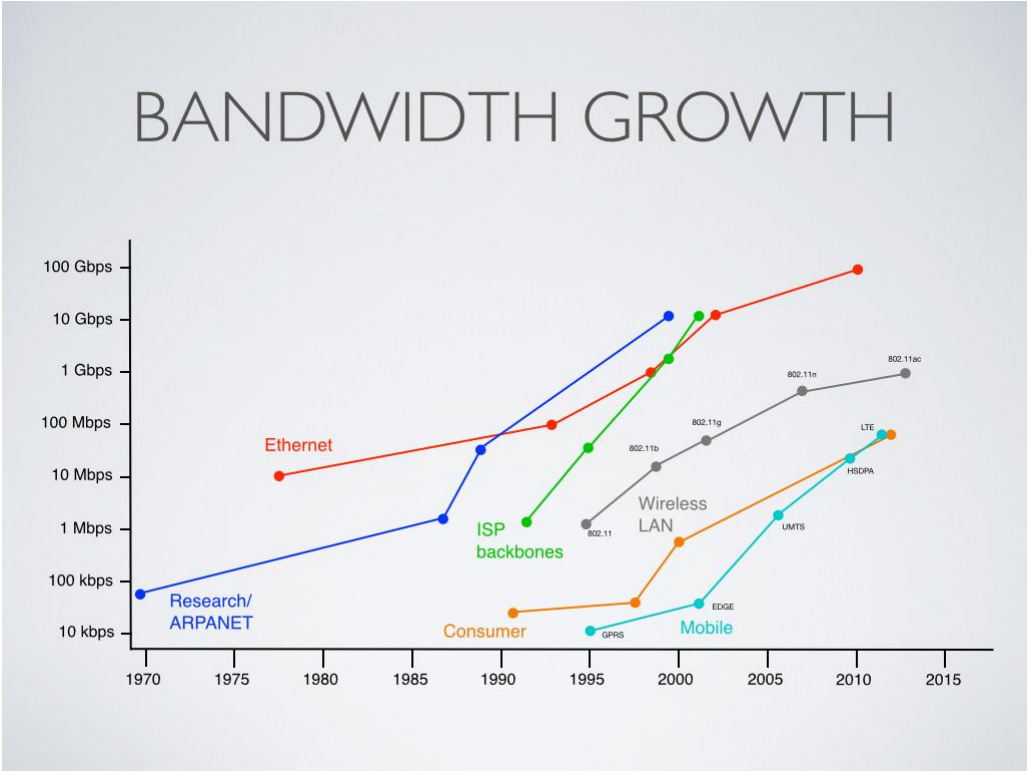
Interaction was limited to exchange of small data due to network limitations

Parallel processing

- In early 1980s it was believed that computer performance could only be improved by producing faster and more powerful processors
- ***Parallel processing*** changed this idea: multiple processors (installed on the same PC or on different PCs) working together to solve a single (big) task that could not be executed on a single one
- The task must be designed to be parallel, i.e. composed by a set of subtasks that could be executed in parallel

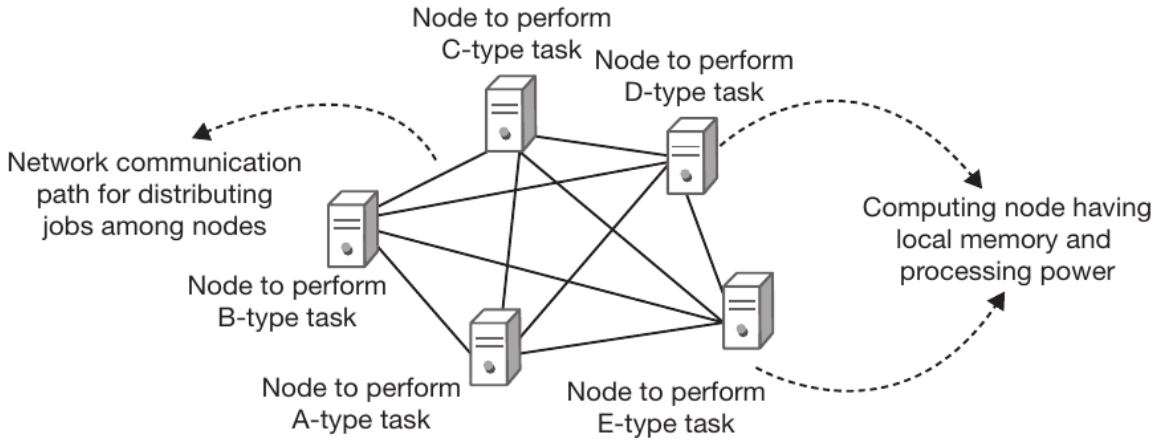
Faster Network Communication

- **<u>During 1980s advancements in network communication improved significantly data transfer rates (up to 100 Mbps for LANs and up to 64 Kbps in WANs)</u>**
- Such advancements enabled ***distributed computing systems***: *a collection of processors interconnected via communication network* to execute a complex task, impossible to be executed on a single PC



Distributed computing systems

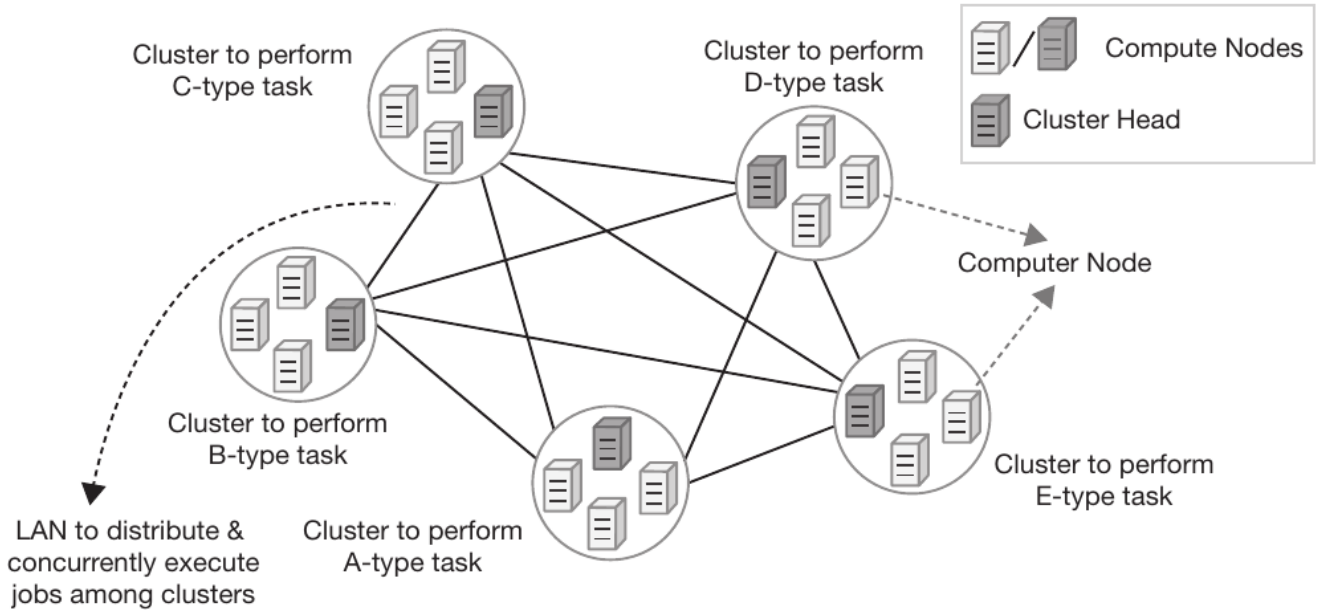
- Applications are programmed to be executed in a distributed manner as a composition of different tasks
- Each system has its own local memory and processing power
- Nodes communicate to exchange data or results



Cluster computing

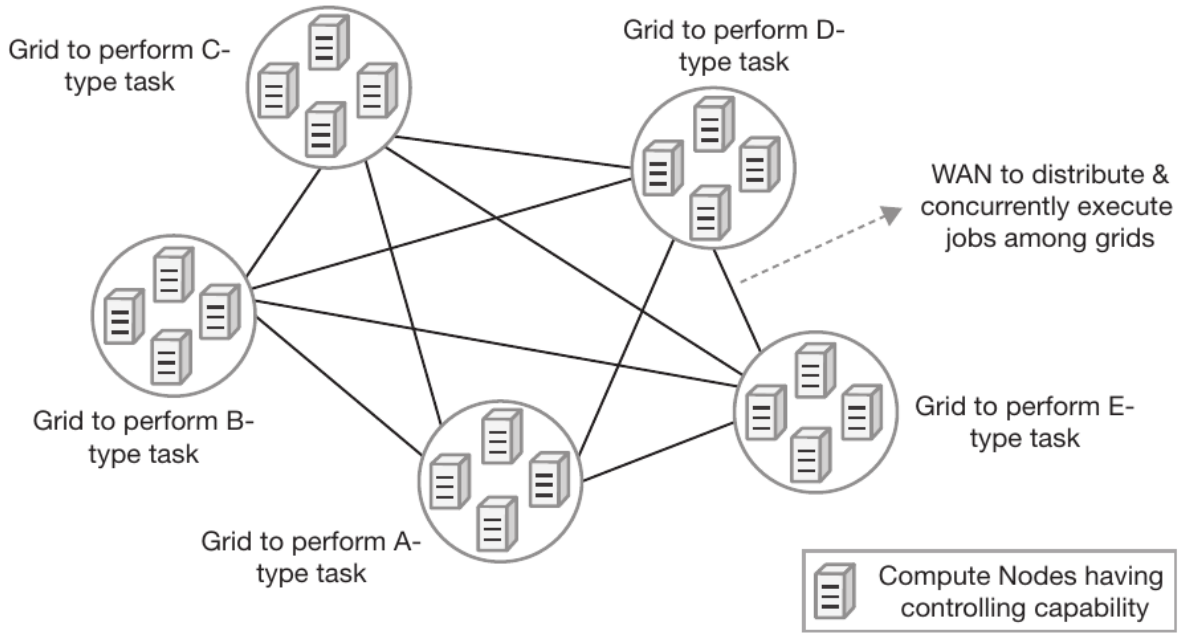
- The next natural step was the creation of clusters: groups of multiple nodes (computers) all connected to the same LAN to perform the same (or similar) tasks
- Computers are clustered together for a more efficient computing (less communication overhead) and to improve reliability.

- Pools of homogeneous computing systems were deployed to have more processing power
- Among the computers of a cluster, one node was selected as cluster head to control the cluster and distribute the tasks



Grid Computing

- The cluster head represented a single point of failure
- To mitigate this risk in early 90s a new architecture was introduced, **grid computing**: *a pool of resources in which the control functionalities were implemented in a decentralized manner (not in the cluster head)*
- Computing nodes could be deployed in the same network or could be deployed in different areas (even under different administrative domains)



Not enough

- Grid computing provided many advantages: *a scalable architecture that could offer high performance computing for complex tasks*
- *It had many drawbacks*:
  - Real-time scaling was not possible: more computing power/storage -> more nodes -> change in the architecture
  - The system was not fault tolerant: the failure of a node results in the failure of a task
  - Heterogeneous hardware required code adaptation
- Starting from the late 90s a new set of technologies matured, paving the way for cloud computing

Hardware Virtualization

- The heterogeneous natures of computing systems architectures posed a challenge to application (software) portability
- The execution of software on different architecture without changes required to decouple software systems from the underlying physical resources of computers
- This would allow to deal with hardware diversity but also to achieve real time scalability
- For this reason, virtualization techniques were defined: a layer of software over the hardware system that could simulate the whole physical system environment

⚠ **We will see that extensively, don't worry!**

Web Based Technologies

- As computing technologies were evolving rapidly, users and business were becoming more dependent on computer systems
- In addition to enabling grid computing to empower the development of large computing systems, faster networks allowed users from different geographical locations to collaborate in almost real time
- The World Wide Web arose as the killer application to spread information and enable collaboration
- Web 2.0 appeared after 2002 as an evolution of Web 1.0, in which user generated content was central

Service Oriented Architecture

- As the number of IT systems and their size increased, a new model for software development was required
- Service Oriented Architecture (SOA) is a development method in which applications are developed by leveraging software components (software services) interacting each other
- Each service is an independent entity that communicate with the others via messages
- Systems implemented using SOA remain flexible to changes

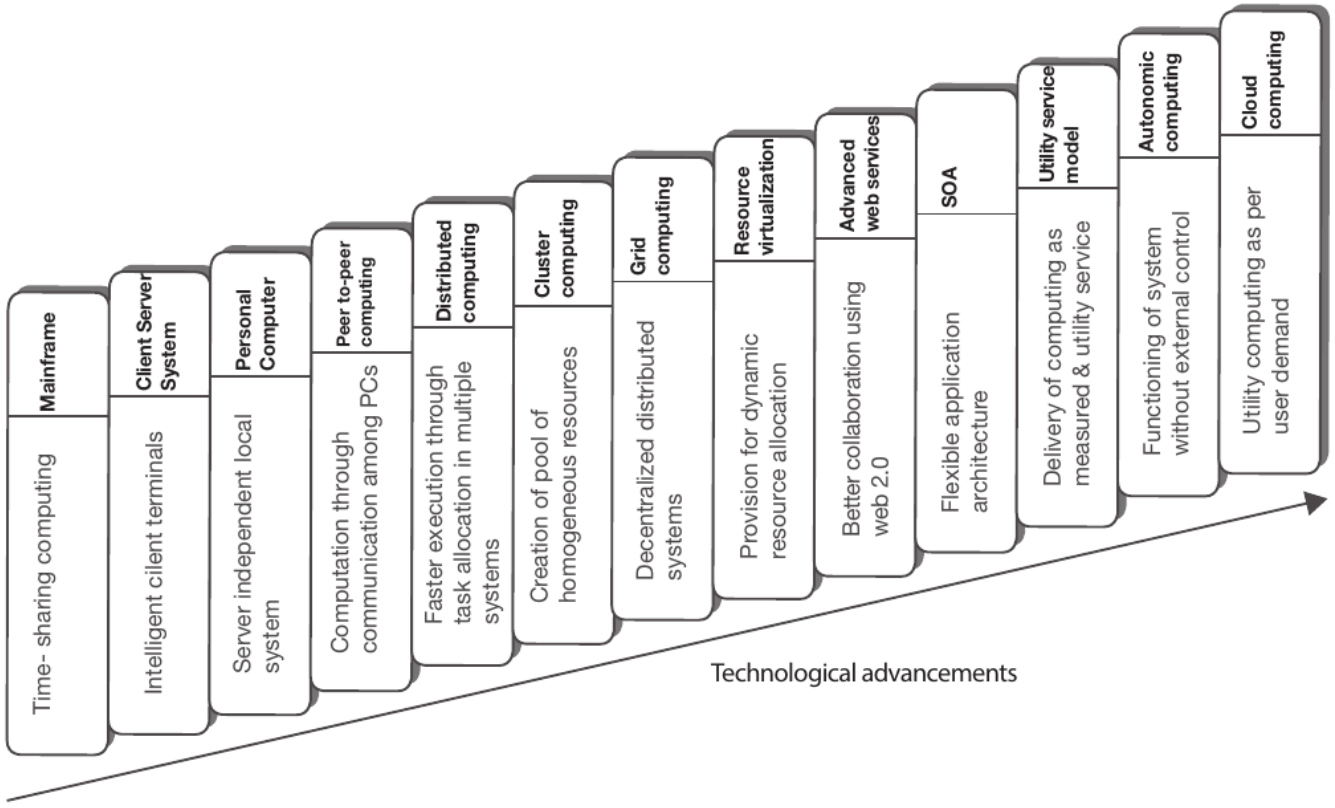
We will see that, don’t worry!

Utility Computing Model

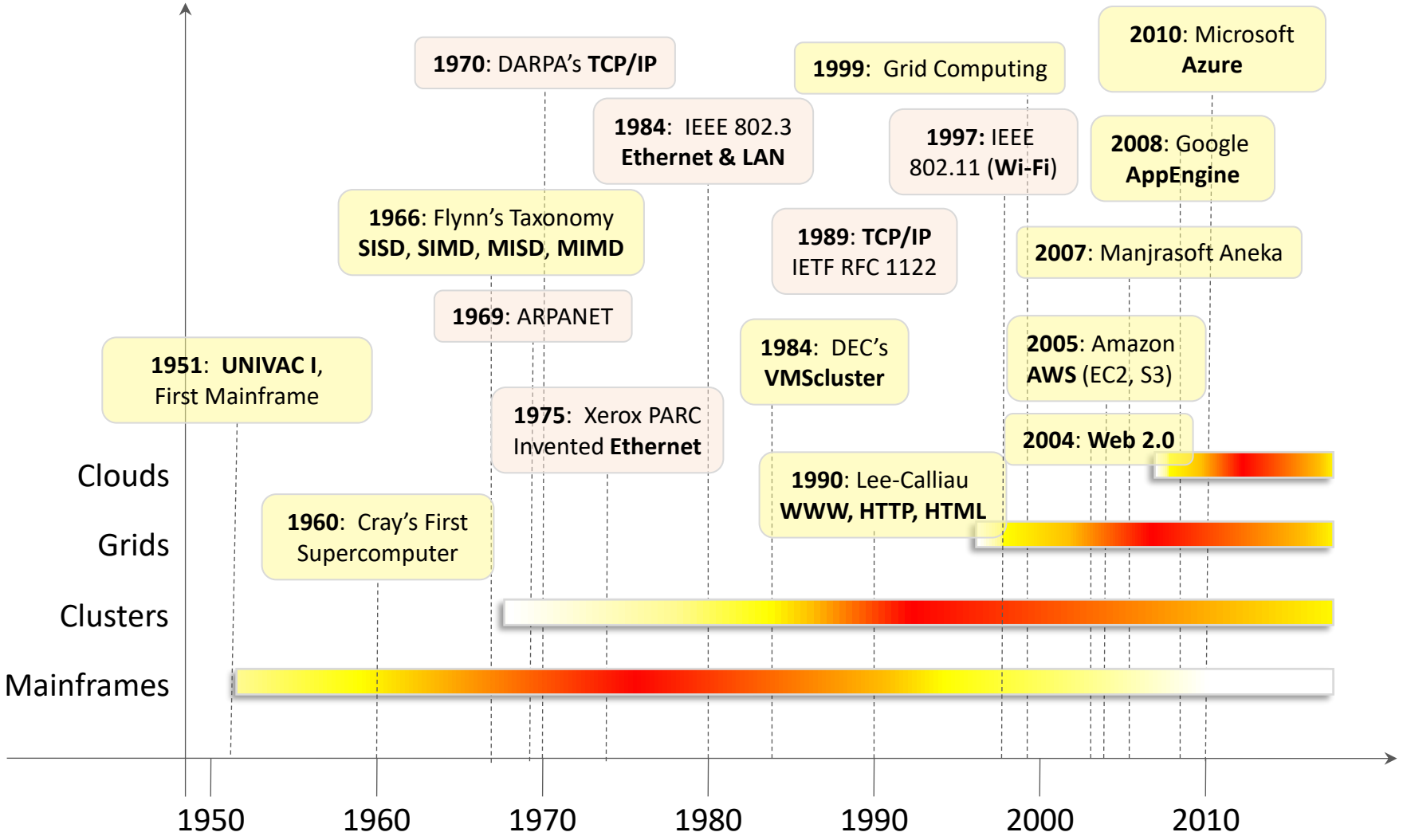
- Computing era reached a point in which we had:
  - Scalable computing infrastructure made of heterogeneous resources
  - Distributed computing environments empowered by high bandwidth networks
  - Collaborative environment across the world via web services
  - Flexible applications via SOA
- Utility Computing: delivering of computing resources as utility, following a pay per use model and on demand
- Crucial **enabler** for this model was hardware virtualization

Cloud Computing … eventually

- Utility computing become cloud computing
- Each one was essential, from networking advancements to novel software development methodologies



Timeline


1970: DARPA's TCP/IP
1999: Grid Computing
2010: Microsoft Azure
1984: IEEE 802.3 Ethernet & LAN
1997: IEEE 802.11 (Wi-Fi)
2008: Google AppEngine
1966: Flynn's Taxonomy SISD, SIMD, MISD, MIMD
1989: TCP/IP IETF RFC 1122
2007: Manjrasoft Aneka
1969: ARPANET
1984: DEC's VMScluster
2005: Amazon AWS (EC2, S3)
1951: UNIVAC I, First Mainframe
1975: Xerox PARC Invented Ethernet
2004: Web 2.0
1990: Lee-Calliau WWW, HTTP, HTML
1960: Cray's First Supercomputer
Clouds
Grids
Clusters
Mainframes
1950
1960
1970
1980
1990
2000
2010

Cloud computing today

- The first commercial initiative based on cloud computing concepts was from salesforce.com in 1999
- The first large scale cloud computing service was commercialized by Amazon in 2006 that lunched its Elastic Cloud Computing (EC2) and Simple Storage Service (S3)
- Soon after other big players entered in the market with similar services, e.g. Microsoft in 2009
- The term Cloud computing appeared with its present meaning for the first time in 2006, used by Google CEO Eric Schmidt in a conference

Cloud computing in numbers

- The total expenditure worldwide reached a figure of $210 billion at the end of 2020
- Demand for cloud services grows by 18% in 2019
- 90% of companies use some type of cloud service
- 80% of enterprises use Amazon Web Services as their primary cloud platform
- There were 3.6 billion cloud users in 2018
- The positive impact of cloud technology is almost instantaneous. 80% of companies report operation improvements within the first few months of adopting the tech