UNIVERSITY OF MARYLAND

UMD DATA605 - Big Data Systems

11.2: Technologies Enabling Cloud Computing

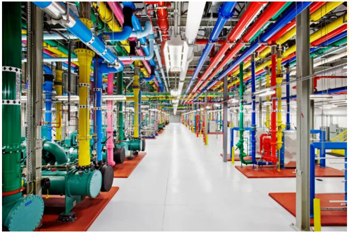
- **Instructor**: Dr. GP Saggese, gsaggese@umd.edu

SCIENCE ACADEMY

- ***Data centers***
- Virtualization
- Programming frameworks
- Challenges and opportunities

Data Centers: Capex

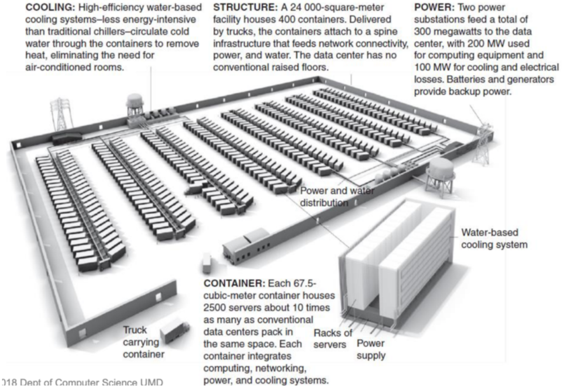
- Data centers enable cloud computing
- Large companies (e.g., AWS, Apple, Google, Facebook) build data centers globally
- Data center costs around 1B USD to build
  - Capex: Computing, memory, storage, networking
  - Prices dropping
  - Size is increasing



Data Centers: Opex

- Powering equipment cost
  - Focus on energy-efficient computing
- High cooling cost
  - Vent placement is key to manage thermal hotspots
  - PUE (Power Usage Effectiveness)
    - Some power converted into computation
    - Rest is overhead
  - Hard to optimize in small data centers
    - Ideal PUE is 1
    - Current is 1.07-1.22
    - May lead to large data centers soon
- Research on energy-saving
- Data centers built
  - Close to cheap energy sources
  - In cold climate

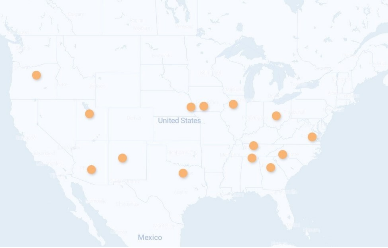
(Modular) Data Centers



Meta

- Global scale of data centers
  - 29 dedicated Meta data centers
  - Hyperscale and AI-optimized facilities
- Investment
  - $600 billion in U.S. infrastructure by 2028
  - El Paso data center: $1.5 billion
  - Louisiana campus: $10 billion for 9 buildings
  - Power infrastructure upgrades: over $3 billion

Data Center	Online	Buildings	SqFt (m)	Investment ($bn)
Dekalb, Illinois	2022	2	0.9	$0.8
Altoona, Iowa	2014	10	4.1	$2.0
Papillion (Sarpy), Nebraska	2019	8	3.6	$1.5
New Albany, Ohio	2020	5	2.5	$1.0
Huntsville, Alabama	2021	4	2.5	$1.0
Newton, Georgia	2023	5	2.5	$1.0
Forest City, North Carolina	2012	4	1.3	$0.8
Gallatin, Tennessee	2023	2	1.0	$0.8
Henrico, Virginia	2020	7	2.5	$1.0
Mesa, Arizona	Q4 2023	2	1.0	$0.8
Los Lunas, New Mexico	2019	6	2.8	$1.0
Fort Worth, Texas	2017	5	2.6	$1.5
Prineville, Oregon	2011	11	4.6	$2.0
Eagle Mountain, Utah	2021	5	2.4	$1.0
Odense, Denmark	2019	2	0.9	$1.6
Clonee, Ireland	2018	3	1.6	$0.4
Luleå, Sweden	2013	3	1.0	$1.0
Tanjong Kling, Singapore	2022	1	1.8	$1.0
Total		**85**	**39.6**	**$20.1**



Amazon Web Services

- AWS
  - 2022: 28 geographical regions
  - 2025: 38 regions

SCIENCE ACADEMY

Amazon Web Services (EC2)

- Widely used solution for cloud computing
  - Many alternatives to suit your needs
  - Prices are low due to competition
  - Current on-demand pricing

Small Instance – default*

1.7 GB memory
1 EC2 Compute Unit (1 virtual core with 1 EC2 Compute Unit)
160 GB instance storage
32-bit platform
I/O Performance: Moderate
API name: m1.small

Large Instance

7.5 GB memory
4 EC2 Compute Units (2 virtual cores with 2 EC2 Compute Units each)
850 GB instance storage
64-bit platform
I/O Performance: High
API name: m1.large

Extra Large Instance

15 GB memory
8 EC2 Compute Units (4 virtual cores with 2 EC2 Compute Units each)
1,690 GB instance storage
64-bit platform
I/O Performance: High
API name: m1.xlarge

Amazon S3

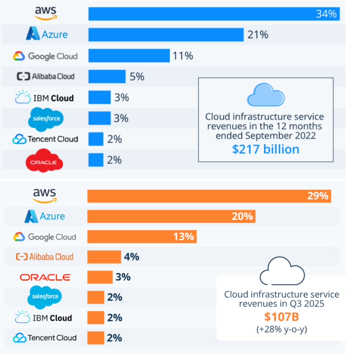
- Amazon storage services (S3 = Simple Storage Solution)
  - Pay for storage you use
- Different tiers for reliability, cost, performance

	Default	**RRS**	**IA**	**Glacier**
Durability	99.999999999%	99.99%	99.999999999%	99.999999999%
Availability	99.99%	99.99%	99.9%	99.99%
Extra Fees	None	None	Retrieval	Retrieval
Real-Time Access?	Yes	Yes	Yes	No (mins/hours)
Frequently Accessed?	Yes	Yes	No	No

Google App Engine

- **Google Compute Engine** (IaaS)
  - Competes with AWS EC2
- **Google Infrastructure** (PaaS)
  - Run Docker containers on Google resources
  - Managed services (e.g., databases)
- **Google Docs** (SaaS)
  - Word processor, spreadsheet, presentations in the cloud

- Google Cloud's **compute market share**
  - Built software data centers before Amazon
  - Invented cloud technologies (e.g., Google File System, MapReduce, BigTable)
  - Market share 3x smaller than AWS
  - Issues:
    - Developer/customer unfriendliness
    - Lack of commitment (Killed by Google)
    - Poor customer service



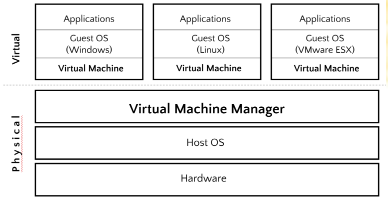
- Data centers
- ***Virtualization***
- Programming frameworks
- Challenges and opportunities

Virtualization

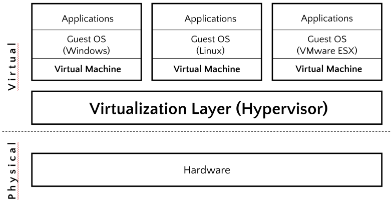
- **Virtual machines have been around for a long time**
  - Processors had support since 1980s
  - In 2000s efficient enough for cloud computing
- **Basic idea of cloud computing**
  - Run virtual machines on servers and sell time on them
  - E.g., AWS, Microsoft Azure, Google Cloud
- **Many advantages**
  - *Security*: virtual machines have almost impenetrable boundary
  - *Multi-tenancy*: multiple VMs on the same server
  - *Efficiency*: replace many underpowered machines with fewer high-powered machines

Desktop vs Server Virtualization

- **Desktop virtualization**
  - VMWare, Xen, VirtualBox
  - Run on host OS
  - Hypervisor/VM supports guest OS
- **Server virtualization**
  - Run hypervisor on hardware
  - Ideal for server farms, cloud computing
  - Amazon used Xen on RedHat
    - Now it uses AWS Nitro
- **Performance are tricky**
  - Hard to reason about performance
  - Identical VMs may deliver different performance
  - Multi-tenancy, different hardware
  - "Bare-metal" compute to improve performance



Consumer / desktop virtualization

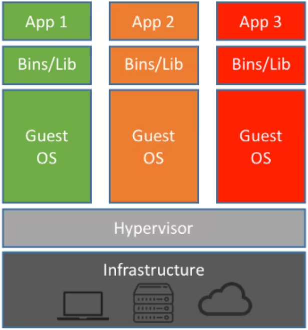


Server virtualization

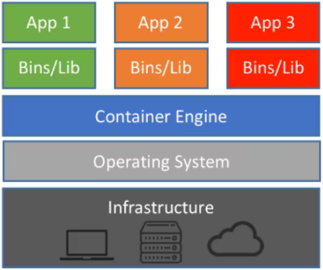
SCIENCE ACADEMY

Docker

- Package all dependencies in a single object
- **Advantages**
  - Containers are fast and portable
  - Reduce virtualization overhead
  - All containers run on a single host
  - Reduce OS licensing cost / maintenance overhead



Machine Virtualization



Containers

SC ACADEMY

- Data centers
- Virtualization
- ***Programming frameworks***
- Challenges and opportunities

Programming Frameworks

- **Programming frameworks** emerged to:
  - Scale out workloads
  - Distribute work over thousands of machines
- **Parallel approach** has existed for a long time to program clusters
  - Challenging for programmers
    - Parallelize application
    - Distribute data
    - Handle failures
    - Debugging
    - Race conditions/Heisenbugs
- **The difference is the user interface**
  - Google developed MapReduce starting a new era
    - Hadoop, Spark
  - AWS services

MapReduce Framework

- Provide a **restricted but powerful abstraction** for programming distributed workloads
- **Separation of responsibilities**
  - *Programmers*
    - Write functions `map` and `reduce`
    - Perform arbitrary computations on input data within structure
  - *Framework* handles
    - Task scheduling
    - Fault tolerance

Other Programming Frameworks

- Many programming frameworks for different applications to address MapReduce limitations
  - **High-performance Computing (HPC) Systems**
    - Cluster of supercomputers
    - E.g., GridRPC, MPI
    - More expressive, efficient
  - **Spark**
    - Based on Resilient Distributed Data (RDD)
    - In-memory, efficient
  - **Apache Storm, Spark Streaming**
    - Handle real-time streaming data
  - **Giraph, GraphLab, GraphX**
    - Graph processing systems
  - **Apache Hive**
    - SQL-like interface on Hadoop/HDFS
  - **Apache HBase**
    - NoSQL column-oriented DB
    - Random read/write large tables on Hadoop/HDFS
    - Modeled after Google BigTable

- Data centers
- Virtualization
- Programming frameworks
- ***Challenges and opportunities***

Cloud Benefits (1/2)

- **Lower-cost, light end-devices**
  - Offload compute/storage to cloud $\rightarrow$ PCs/laptops need less memory/disk/CPU
  - Useful for thin clients, low-cost devices, legacy hardware
- **Scalability and elastic storage**
  - Unlimited storage and dynamic scalability for data/workloads
  - No need to buy or pre-provision resources
- **Anywhere access and device independence**
  - Work from any device (laptop, tablet, phone) with Internet
  - Documents, apps, data follow user: seamless cross-device continuity

Cloud Benefits (2/2)

- **Cloud-native software and SaaS model**
  - Access full-featured applications without installation or license
  - Automatic updates and patching, always-on latest versions
- **Improved collaboration and version control**
  - Real-time, multi-user editing and sharing from any location
  - Built-in revision history and conflict avoidance for shared documents
- **Faster development and deployment cycles**
  - Containerization, serverless functions, microservices enable rapid provisioning
  - Ideal for modern workloads (e.g., AI/ML, analytics, distributed apps)

Modern Opportunities

- **Cloud-native support for AI/ML workloads**
  - Providers offer specialized hardware (GPUs, TPUs) for training/serving models
  - Enables scaling without owning expensive hardware
- **Multi-cloud and hybrid-cloud strategies to boost resilience**
  - Organizations adopt multicloud fallback after major outages
- **Improved data sovereignty and regulatory compliance**
  - New regulations (e.g. EU Data Act of 2025) increase portability and reduce vendor lock-in
  - "Sovereign cloud" options for jurisdiction and legal control concerns

Cloud Limitations

- **Cloud provider dependency and vendor lock-in**
  - Reliance on major providers limits flexibility and competition
  - Switching providers is complex and costly
- **Security, privacy, and data-ownership issues**
  - Data stored off-site may be subject to foreign access laws
  - Sensitive workloads require encryption, strong data governance, or on-premises solutions
- **Outages and service disruptions**
  - Major outages continue
  - Even top providers are not immune; "design for failure" remains essential
- **Latency, bandwidth, and connectivity dependence**
  - Cloud services require stable, fast Internet (problematic in regions with spotty connectivity)
  - High-latency workloads (e.g., real-time gaming, low-latency HPC, real-time control) may suffer

Post-2025 Challenges

- **Increasing regulatory burden and data-sovereignty demands**
  - Regulations like EU Data Act require cloud providers to support data portability, transparency, restricted cross-border access
  - Compliance adds complexity for companies across jurisdictions
- **Cloud sprawl and rising costs**
  - Multicloud and hybrid setups complicate cost management, egress charges, licensing
  - Resource usage can increase without careful monitoring
- **Feature and performance limitations for specialized workloads**
  - Desktop-level apps and advanced tools often outperform web/cloud versions
  - Compute-intensive or latency-sensitive tasks may remain sub-optimal in the cloud