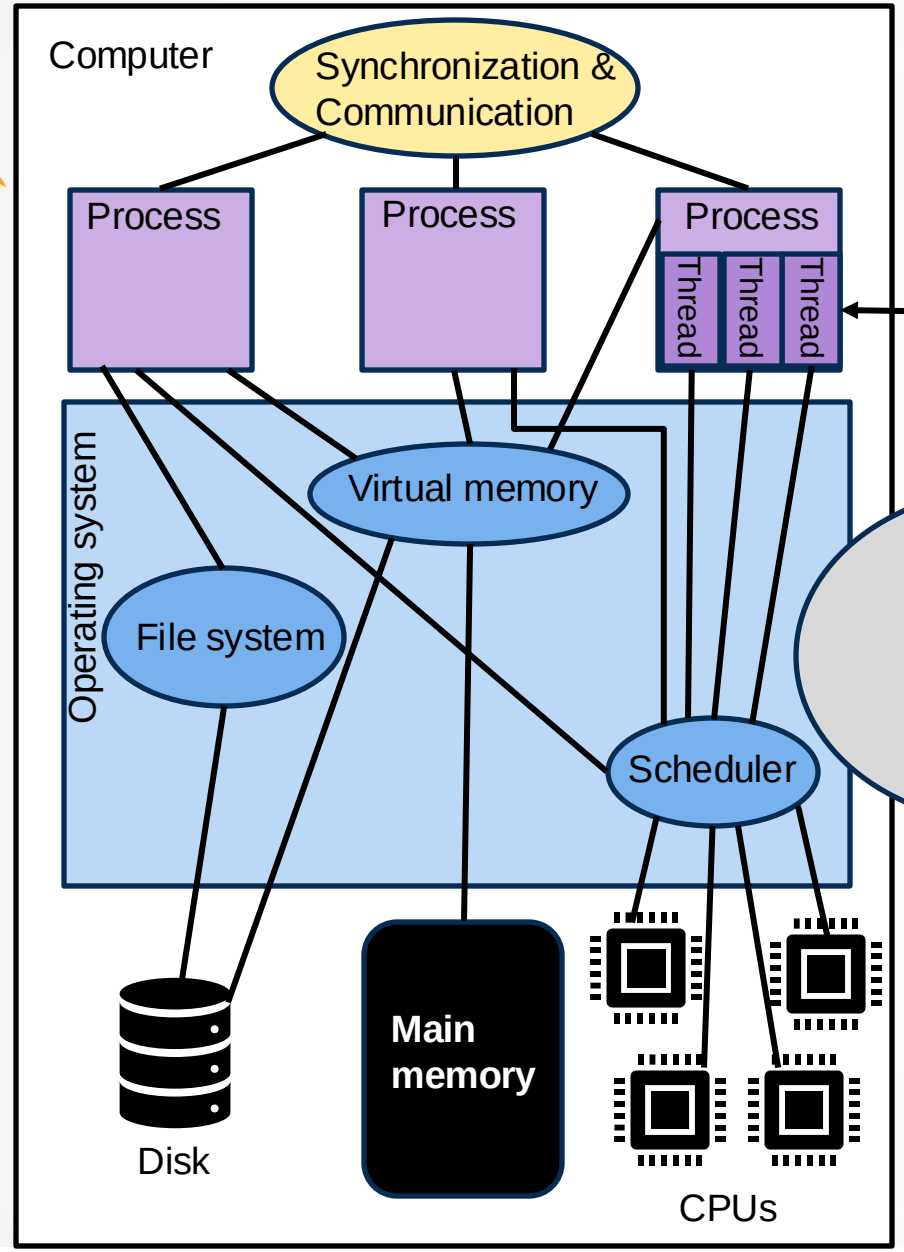
COMPUTING PLATFORMS

Cloud Computing

OVERVIEW OF LECTURE

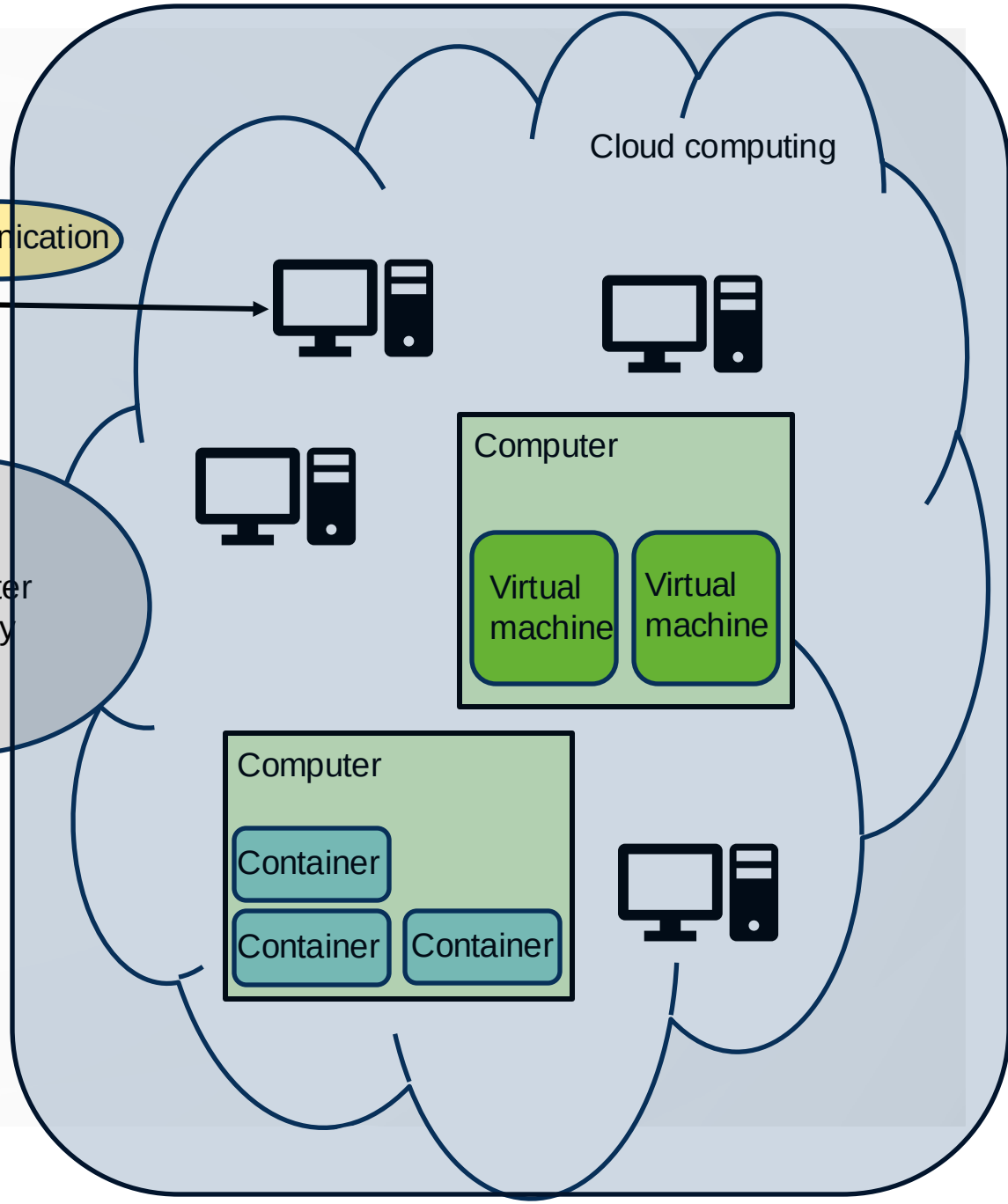
- Cloud computing overview
  - Definition and service models (IaaS, PaaS, SaaS)
  - Deployment models (public, private, hybrid)
- Cloud infrastructure and management
  - Managing cloud resources
  - Cost-benefit analysis
- Cloud security
  - Challenges and solutions
- Future trends and innovations
  - Emerging technologies in cloud computing
- No technical details, mainly high-level issues

CONTENTS OF THE COURSE



Communication

Computer security



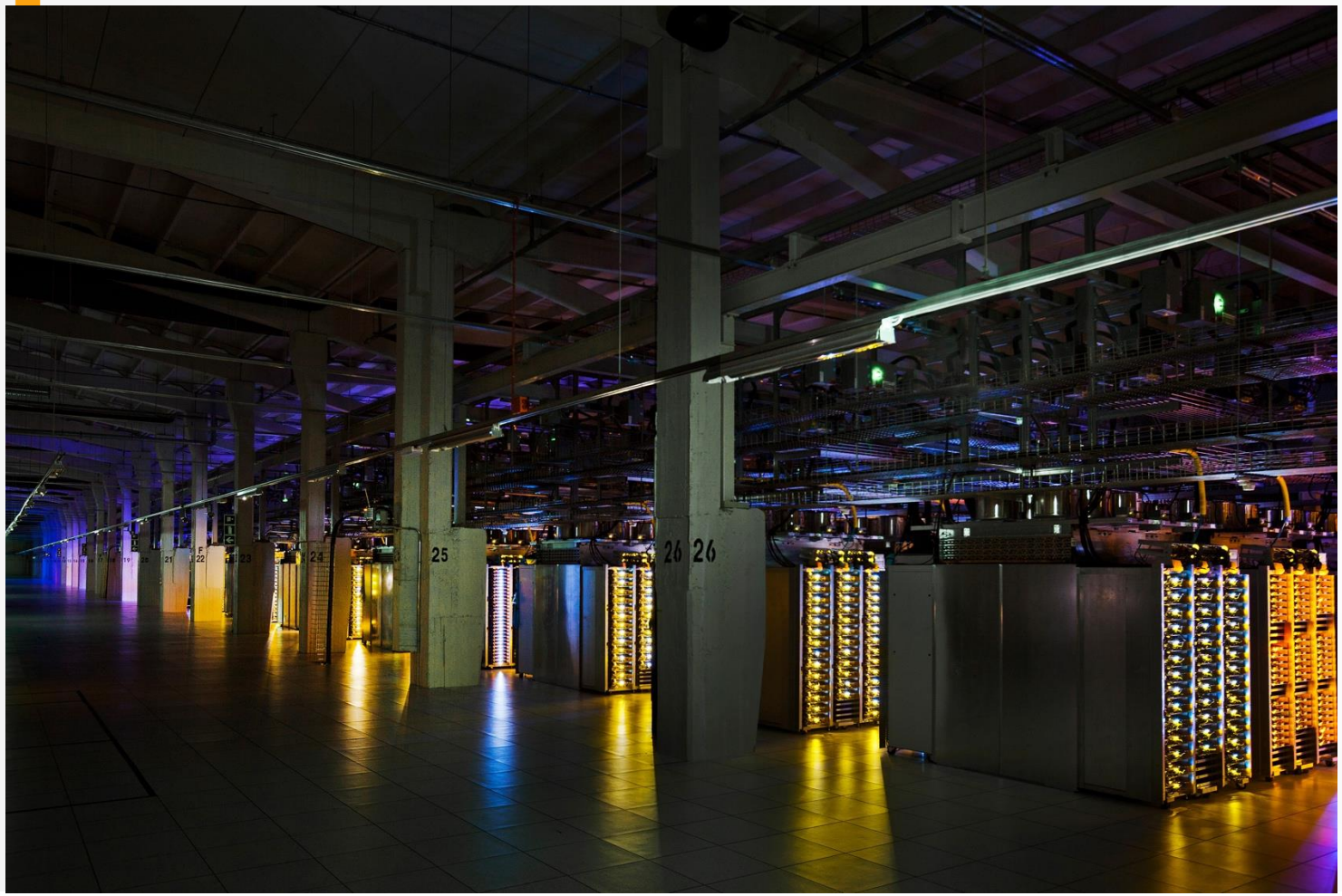
DEFINITION OF CLOUD COMPUTING

- **Delivery of computing services over the internet including servers, storage, databases, networking, …**
- On-demand access to shared computing resources; no active management by the user
- Resources are pooled to serve multiple consumers, using a multi-tenant model
- Services are available over the network
- Elastically provisioned and released to scale rapidly outward and inward with demand
- Automatically control and optimize resource use by leveraging a metering capability
- **Service Models**: Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS)
- **Deployment Models**: Public cloud, private cloud, hybrid cloud, and community cloud.

WHAT DOES THE CLOUD LOOK LIKE?

Google’s datacenter
in Hamina
Source: www.google.com

WHAT DOES THE CLOUD LOOK LIKE?



Google’s datacenter
in Hamina
Source: www.google.com

BRIEF HISTORY OF CLOUD COMPUTING

THE EMERGENCE OF CLOUD COMPUTING (2010-2012)

- Earliest ideas from 1996 (Compaq), in 2003 beginning of Amazon Web Services (AWS)
- 2006: Launch of Amazon Elastic Compute Cloud
- 2010: Mainstream adoption
  - Major Cloud Providers: Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform
- 2011: Proliferation of Software as a Service (SaaS)
  - Applications like Salesforce, Google Apps
- 2012: Improved Infrastructure as a Service (IaaS) and Platform as a Service (PaaS) offerings
- **Hybrid Cloud concept**: Emergence of the hybrid cloud model, combining private and public clouds
- **Big Data and the cloud**: Increasing use of cloud computing for big data analytics
- **Cloud security focus**: Growing emphasis on cloud security and compliance standards

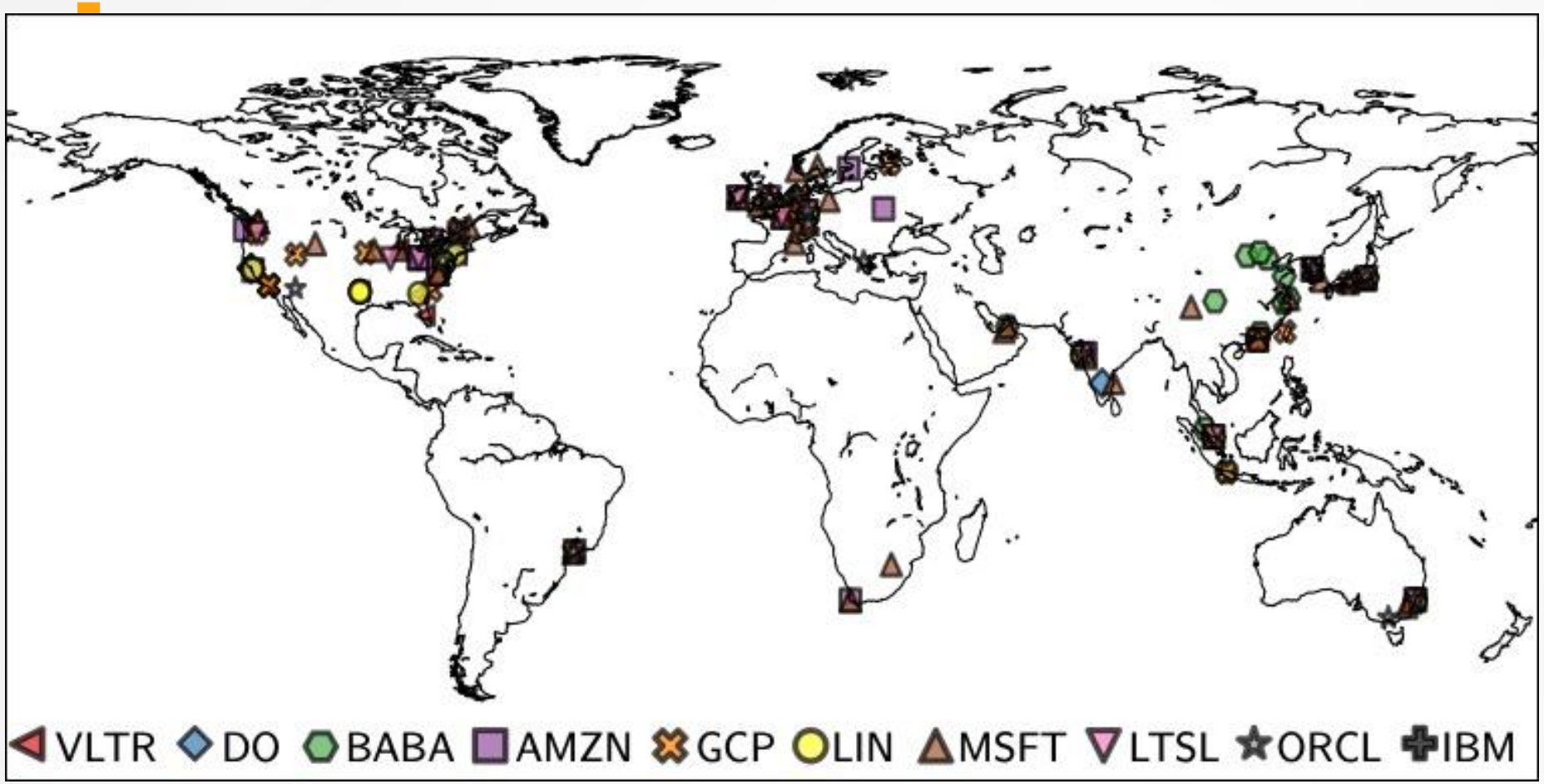
ADVANCEMENTS AND EXPANSION (2013-2016)

- 2013: Cloud becomes the ”New Normal”: More enterprises adopt cloud-first strategies
- Expansion of AWS Services: Introduction of innovative services like AWS Lambda for serverless computing
- 2014: Containerization and Docker: Docker popularizes container technology, impacting cloud deployment models
- 2015: Cloud and mobile computing: Integration of cloud services with mobile applications
- Microsoft Azure Growth: Enhancements and adoption of Microsoft Azure
- 2016: The rise of AI and ML in the cloud: Cloud platforms begin offering AI and ML services
- Internet of Things (IoT) and cloud: IoT devices increasingly rely on cloud for data processing and storage

MATURATION AND NEW HORIZONS (2017-2020)

- 2017: Multi-Cloud strategies: Organizations start adopting multi-cloud approaches for flexibility and risk management
- Kubernetes and orchestration: Widespread adoption of Kubernetes for container orchestration
- 2018: Edge computing emergence: Edge computing gains prominence for low-latency processing
- Cloud gaming and streaming services: Cloud technology begins powering gaming and streaming platforms
- 2020: Cloud computing and the Covid-19 pandemic: Rapid acceleration in cloud adoption due to remote work and online collaboration needs
- Sustainability in cloud computing: Increased focus on sustainable, energy-efficient cloud services

EXTENT OF CLOUD TODAY



Subset of global cloud datacenters

From The Kang Dang et al. Cloudy with a Chance of Short RTTs Analyzing Cloud Connectivity in the Internet, ACM IMC 2021

CLOUD SERVICE MODELS

CLOUD SERVICE MODELS

- Three common service models:
  - Infrastructure as a Service (IaaS)
  - Platform as a Service (PaaS)
  - Software as a Service (SaaS)
- Each provides a different kind of a solution for different use cases
- Whole bunch of other “X as a Service” exist also

INFRASTRUCTURE AS A SERVICE (IAAS)

- IaaS provides virtualized computing resources over the internet
- Includes servers, storage, networking, and virtualization
- Users have control over their infrastructure without managing physical hardware
- Resources can be scaled up or down based on demand
- Typically operates on a pay-as-you-go pricing model
- Web hosting, storage and backup, web apps, high-performance computing
- **Popular providers**: Amazon Web Services (AWS), Microsoft Azure, Google Cloud Platform
- **Target users**: Suitable for businesses that want to avoid the cost and complexity of purchasing and managing their physical servers

PLATFORM AS A SERVICE (PAAS)

- Platform allowing customers to develop, run, and manage applications without the complexity of building and maintaining infrastructure
- Operating systems, middleware, development tools, database management systems
- Create applications without worrying about underlying infrastructure
- Tools for software development, such as source code editors and version management
- Allows for customization while abstracting away the hardware-level details
- **Examples**: Google App Engine, Microsoft Azure App Services, Heroku
- **Benefits**: Reduces the complexity of the software development process
- **Target users**: Developers and companies focusing on software development and deployment

SOFTWARE AS A SERVICE (SAAS)

- SaaS delivers software applications over the internet, on a subscription basis
- Accessible from various devices over the internet with a web browser
- Users don’t need to manage, install, or upgrade software
- Services are scalable with options for subscription tiers
- Often includes data analytics and reporting capabilities
- **Examples**: Google Workspace, Salesforce, Microsoft Office 365, Dropbox
- Providers manage the security, compliance, and maintenance
- **Target users**: Ideal for businesses seeking software solutions without the need for extensive IT infrastructure

SOME OTHER ”X AS A SERVICE” MODELS

- **Desktop as a Service (DaaS)**:
  - Provides virtual desktops hosted on remote servers
  - Offers flexibility and cost savings on hardware and software maintenance
- **Database as a Service (DBaaS)**:
  - Offers database management capabilities without the need to set up physical hardware, install software, or configure for performance
  - Examples include Amazon RDS and Microsoft Azure SQL Database
- **Function as a Service (FaaS)**:
  - A form of serverless computing where developers can execute code in response to events without the complexity of building and maintaining the infrastructure
  - Examples include AWS Lambda, Azure Functions, and Google Cloud Functions

INTRODUCTION TO SERVERLESS COMPUTING

- A cloud computing model where the cloud provider manages the infrastructure
- Developers write and deploy code without worrying about the underlying infrastructure
- Functions are triggered by specific events or requests
- Automatically scales up or down based on demand
- Billing based on the actual amount of resources consumed by applications, as opposed to pre-purchased units of capacity
- Functions are stateless, and the execution environment is ephemeral
- **Examples of serverless services**: AWS Lambda, Azure Functions, Google Cloud Functions
- **Ideal for microservices**: Simplifies deployment and management of microservices

HOW SERVERLESS COMPUTING WORKS

- Developers deploy their code to a serverless platform
- No need to provision, maintain, or administer servers
- Code runs in stateless containers that are event-triggered
- Automatically scales with the number of executions
- **Load balancing**: Handled by the cloud provider, distributing incoming requests across multiple instances
- Often used in conjunction with other cloud services like databases, IoT, and analytics
- **Use cases**: Web applications, APIs, data processing, and real-time file processing
- **Limitations**: Timeouts, state management, and vendor lock-in concerns

CLOUD DEPLOYMENT MODELS

COMMON CLOUD DEPLOYMENT MODELS

- 4 common models for deploying cloud computing
  - Public
  - Private
  - Hybrid
  - Community
- Key issues: Who operates it, who can access it, where it can be accessed, …

PUBLIC CLOUD

- Services offered over the public internet and available to anyone who wants to purchase them
  - For example: Amazon Web Services (AWS), Microsoft Azure, Google Cloud Platform
- High scalability due to vast resources available
- Pay-as-you-go pricing model reduces upfront costs
- Managed and maintained by the service providers
- Robust security measures, though shared with other customers
- Adherence to various compliance standards, suitable for non-sensitive data
- **Use cases**: Ideal for small to medium businesses, startups, and for handling sporadic workloads

PRIVATE CLOUD

- Exclusive cloud environment dedicated to a single organization
- Offers greater control and customization options
- Enhanced security and privacy, suitable for sensitive data and regulatory compliance
- Higher initial investment for setup and maintenance
- Can be hosted on-premises or externally by a third-party provider
- Less scalable compared to public clouds but offers more flexibility for customization
- Optimized performance due to dedicated resources
- **Use cases**: Ideal for large organizations with stringent data privacy, security, and regulatory requirements

HYBRID CLOUD

- Combines public and private clouds to allow data and applications to be shared
- Offers balance between scalability and security
- Allows organizations to use the public cloud for high-demand and less-sensitive tasks
- Sensitive data can be kept on a private cloud while leveraging the robust computational resources of a public cloud
- Enhances disaster recovery and business continuity strategies
- Requires compatibility between cloud environments for seamless operations
- Can be more complex to manage due to multiple platforms
- **Use cases**: Suitable for businesses needing a mix of data isolation and scalable resources

COMMUNITY CLOUD

- Shared by several organizations for a specific community with common concerns
- Costs are distributed among the users, making it cost-effective
- Facilitates community-specific collaboration and data sharing
- Tailored security and compliance for the specific community
- Can be managed internally or by a third-party
- Offers a moderate level of scalability based on the community needs
- Customized to serve a specific community, industry, or group with common objectives
- **Use cases**: Ideal for government organizations, educational institutes, or industry-specific applications

CLOUD INFRASTRUCTURE AND MANAGEMENT

SCALABILITY AND ELASTICITY IN CLOUD INFRASTRUCTURE

- **Definition of scalability**: The ability of cloud infrastructure to handle growing workloads by increasing resource capacity.
- **Vertical vs. horizontal scaling**:
  - Scaling up (vertical): Adding more power to existing machines
  - Scaling out (horizontal): Adding more machines
- **Elasticity**: The capability of cloud resources to automatically scale in response to demand
- **Auto-scaling features**: Services for automatic scaling based on predefined rules and metrics
- **Load-adaptive systems**: Designing systems to adapt to workload changes seamlessly
- **Cost implications of scalability**: Balancing performance needs with cost-effective resource use
- **Challenges in scalability**: Addressing potential issues like data consistency, latency, and network bottlenecks
- **Best practices**: Implement effective scalability strategies for optimal performance and reliability

COST MANAGEMENT AND OPTIMIZATION

- **Cloud pricing models**: Pay-as-you-go, reserved instances, and spot pricing
- **Cost visibility and tracking**: Using tools to monitor and report cloud spending
- **Identifying unused/underused resources**: Reduce costs by shutting down idle resources
- **Right-sizing resources**: Matching resource types and sizes to workload requirements
- **Budget alerts and reporting**: Setting up alerts to prevent overspending
- **Cost allocation**: Distributing cloud costs to different departments or projects
- **Optimizing storage costs**: Implementing data lifecycle policies and choosing appropriate storage classes
- **Cloud financial management practices**: Employing a cloud financial management strategy to align cloud spending with business goals

UNDERSTANDING CLOUD PRICING MODELS

- **Pay-as-you-go (PAYG)**
  - Flexibility: Users only pay for the computing resources they use, typically measured per hour or per second
  - No upfront cost: Ideal for businesses seeking flexibility without significant initial investment
  - Adaptable to changing needs: Costs scale up or down based on actual usage, suitable for variable workloads
- **Reserved instances**
  - Cost-effective for predictable usage: Users commit to using a specific amount of resources for a predetermined period (1 to 3 years) to get a lower rate
  - Upfront payment options: Offers various payment options: upfront, partial upfront, or no upfront but commitment
  - Long-term savings: Significant cost savings over PAYG for stable and predictable usage patterns
- **Spot pricing/Spot instances**
  - Cost-efficiency for flexible workloads: Allows users to bid for unused capacity at a potentially lower price
  - Dynamic pricing: Prices fluctuate based on supply and demand
  - Best for non-critical, interruptible tasks: Ideal for workloads that can be interrupted or flexible in terms of timing, such as batch processing or background tasks

COST-BENEFIT ANALYSIS OF CLOUD

- Reduces need for upfront investments in IT infrastructure and hardware
- Go from capital expense (CapEx) to operational expense (OpEx)
- Ability to scale resources up or down based on demand; pay only for what you use
- Reduction in hardware setup and maintenance times
  - Increased productivity and faster time-to-market
- More energy-efficient than traditional data centers
- Cloud platforms offer robust backup and recovery solutions, reducing the cost of data loss and downtime
- Provides access to the latest technologies and innovations without the need for additional investments in new hardware or software
- Better utilization of IT resources, including staff, as cloud providers manage and maintain the cloud infrastructure

PERFORMANCE MONITORING (USER SIDE)

- Ensuring optimal operation and user satisfaction
- **Key metrics to monitor**: CPU, memory usage, disk I/O, network throughput.
- **Real-time monitoring tools**: Using tools like Amazon CloudWatch, Google Stackdriver for live performance tracking.
- **Predictive analytics**: Utilizing analytics to predict and mitigate potential performance issues.
- **Alerting and notification systems**: Configuring alerts for performance anomalies.
- **Log management and analysis**: Aggregating and analyzing logs for performance insights.
- **Application Performance Management (APM)**: Tools specifically designed to monitor application-level performance.
- **Benchmarking and testing**: Regularly benchmarking performance against industry standards and conducting stress tests.

CLOUD SECURITY ISSUES

- **Security challenges in the cloud**: Understanding the unique security risks associated with cloud computing
- **Data security and privacy**: Protect sensitive data from unauthorized access and breaches
- **Identity and Access Management (IAM)**: Controlling who can access what resources in your cloud environment
- **Compliance and legal issues**: Adhering to regulatory standards and legal requirements
- **Encryption and data protection**: Encryption strategies for data at rest and in transit.
- **Threat detection and management**: Identifying and mitigating potential security threats
- **Security best practices**: Establishing robust security protocols and policies
- **Shared responsibility model**: Understanding the division of security responsibilities between cloud providers and users

SECURITY CHALLENGES IN CLOUD COMPUTING

- **Data breaches**: Sensitive data exposure due to misconfiguration, weak encryption, or insider threats
- **Lack of visibility and control**: Challenges in maintaining visibility and control over data and resources in a multi-tenant cloud environment
- **Compliance and regulatory challenges**: Adhering to data protection laws and industry standards across different regions
- **Account hijacking**: Stolen credentials leading to unauthorized access to cloud services
- **Insecure interfaces and APIs**: Risks associated with insecure or poorly designed APIs in cloud services
- **Insider threats**: Risks from malicious insiders within an organization or the cloud service provider
- **Advanced Persistent Threats (APTs)**: Targeted attacks that can infiltrate cloud networks and remain undetected for long periods
- **Distributed Denial of Service (DDoS) attacks**: Sophisticated DDoS attacks targeting cloud resources

DATA SECURITY AND IDENTITY ACCESS MANAGEMENT IN CLOUD COMPUTING

- **Data security in the cloud**: Importance of protecting data stored in cloud services from unauthorized access and breaches
- **Encryption techniques**: Using encryption for data at rest and in transit to ensure data confidentiality
- **Data sovereignty and localization**: Understanding how data residency and sovereignty impact compliance and privacy
- **Identity and Access Management (IAM)**: Tools and strategies to manage digital identities and control access to resources
- **Role-Based Access Control (RBAC)**: Assigning and managing access to resources based on roles within the organization
- **Single Sign-On (SSO) and Multi-Factor Authentication (MFA)**: Enhancing security through robust authentication mechanisms
- **Regular audits and compliance checks**: Performing regular security audits to ensure compliance with policies and standards.

COMPLIANCE, THREAT MANAGEMENT, AND SHARED RESPONSIBILITY

- **Compliance and regulatory standards**: Requirements like GDPR, HIPAA, PCI-DSS
- **Legal implications**: Understanding legal aspects of storing and processing data in the cloud
- **Threat detection and response**: Utilizing cloud-native tools for monitoring, detecting, and responding to security threats
- **Security Information and Event Management (SIEM)**: Integrating SIEM solutions for real-time analysis of security alerts
- **Incident response planning**: Developing and implementing an incident response plan for potential security breaches
- **Shared responsibility model**: Clarifying the security responsibilities of the cloud service provider versus the cloud user
- **Best practices in cloud security**: Implementing security best practices to fortify the cloud environment

BEST PRACTICES IN CLOUD SECURITY

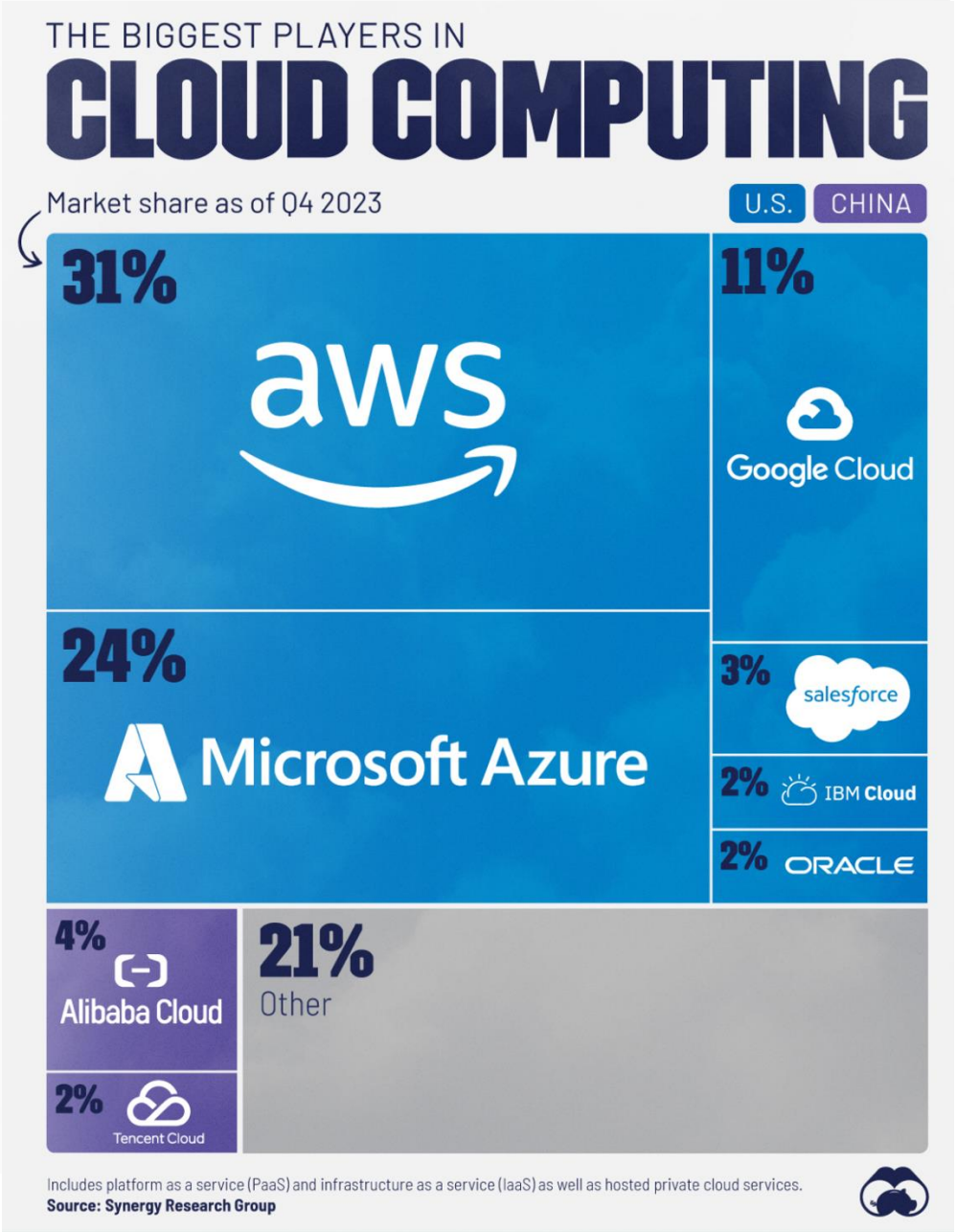
- **Implement strong access control measures**: Use identity and access management (IAM) systems, multi-factor authentication, and role-based access control
- **Data encryption**: Encrypt sensitive data both at rest and in transit
- **Regular security assessments and audits**: Periodic security reviews and compliance audits
- **Use of security tools and services**: Cloud-native security tools for threat detection, prevention, and response.
- **Employee training and awareness**: Training staff on best practices and potential threats
- **Backup and disaster recovery plans**: Implementing robust backup strategies and disaster recovery plans for data and applications
- **Secure APIs and endpoints**: Ensure APIs and endpoints are properly secured and monitored
- **Follow the shared responsibility model**: Clearly understanding and adhering to the security responsibilities shared between the cloud provider and the user

CLOUD NOW AND TOMORROW

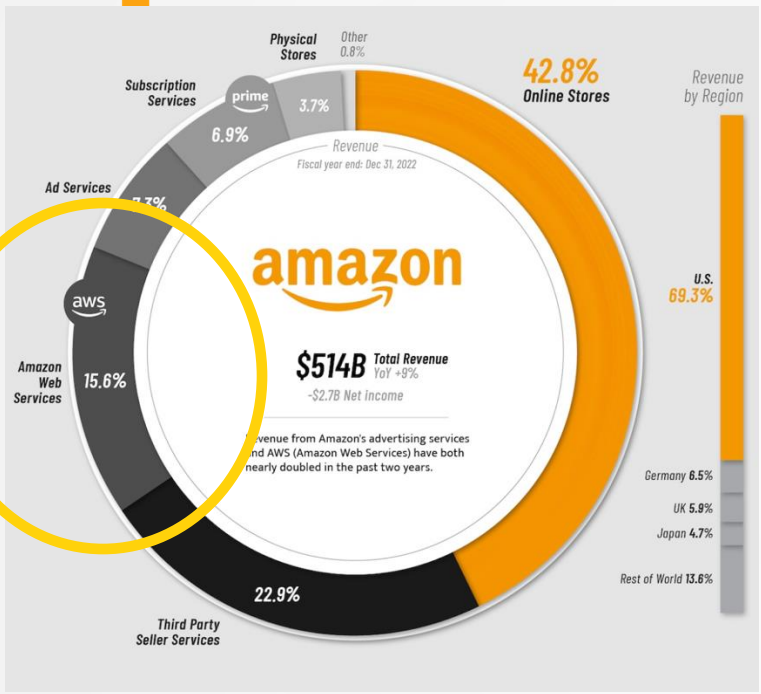
CLOUD MARKET TODAY

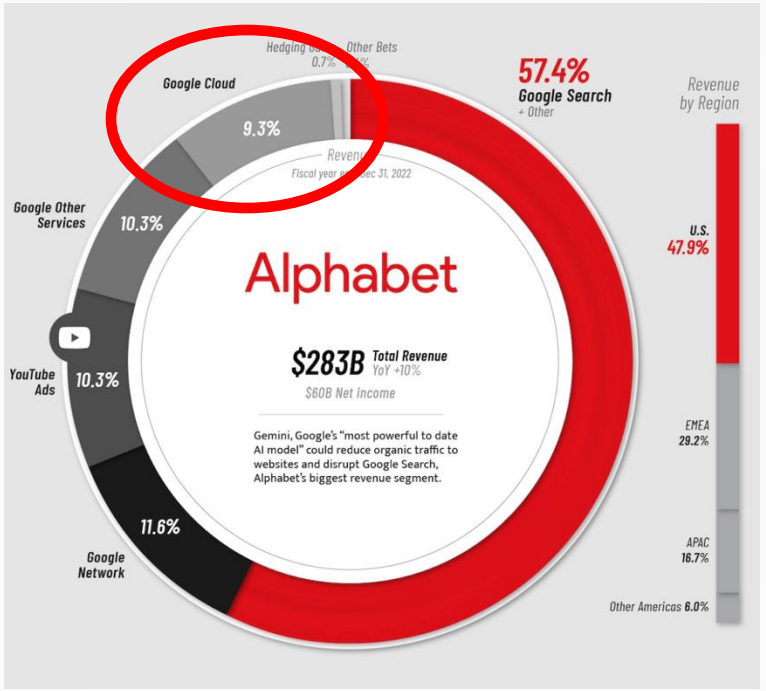
- Hundreds of different cloud providers
- Market dominated by a few players
- No real monopoly for anyone

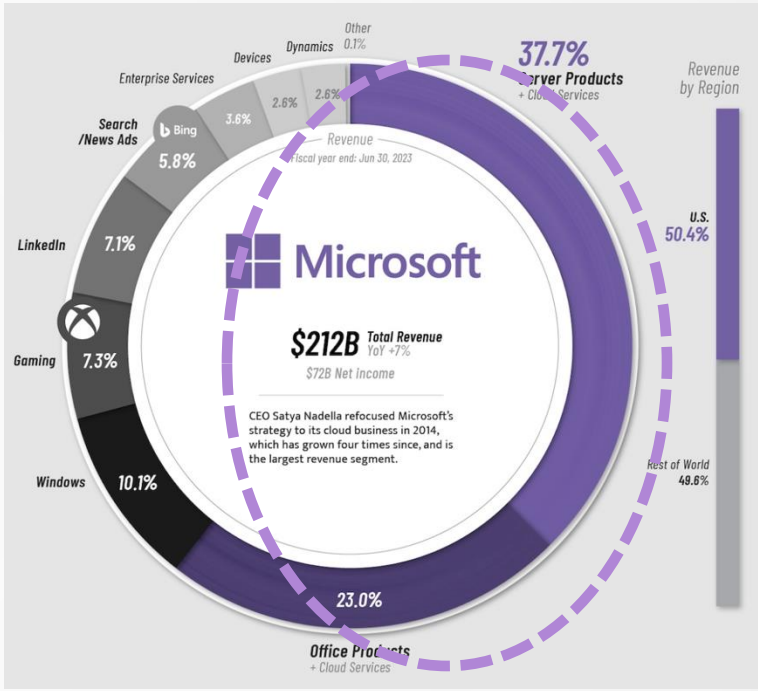
- From: https://www.visualcapitalist.com/worlds-biggest-cloud-computing-service-providers/



CLOUD AS A BUSINESS







Big cloud company revenue streams in 2022
From: https://www.visualcapitalist.com/big-tech-companies-billions/

CURRENT TRENDS IN CLOUD COMPUTING

- **Hybrid and multi-cloud strategies**
- **Serverless computing**
- **AI and ML integration**
  - Integration of AI and ML services for advanced data analytics and automation
- **Containerization and Kubernetes**
- **Edge computing**
  - Expansion of edge computing for faster processing and reduced latency
- **Increased focus on cloud security**
- **Sustainability in cloud computing**
  - Emphasis on green computing and energy-efficient data centers
- **Cloud gaming and streaming services**
  - Growth in cloud-based gaming and streaming platforms

FUTURE DEVELOPMENT DIRECTIONS IN CLOUD COMPUTING

- Quantum computing in the cloud
  - Potential integration of quantum computing resources in the cloud
- 5G and cloud convergence
  - Leveraging 5G technology to enhance cloud applications, particularly in mobile and edge computing.
- Autonomous cloud
  - Self-managing cloud environments using AI and ML for automatic optimization and maintenance.
- Blockchain in the cloud
  - Increased use of blockchain technology for enhancing cloud security, trust, and data integrity.
- Augmented and Virtual Reality (AR/VR)
  - Growth in cloud-based AR and VR applications, particularly in entertainment, training, and education.
- Cloud sovereignty
  - Growing focus on data sovereignty and regional cloud services due to geopolitical and regulatory considerations.

SUMMARY

- Cloud Computing Overview
- Cloud Infrastructure and Management
- Cloud Security
- Future Trends and Innovations