Architectures for the cloud Overview

1. Basic Cloud Definitions

- **NIST Definition**: The National Institute of Standards and Technology (NIST) defines the following essential characteristics of cloud computing:
    - **On-Demand Self-Service**: Users can provision resources like server time and network storage automatically, without human involvement.
    - **Ubiquitous Network Access**: Cloud services are accessible over the network via standard protocols, making them usable across various devices.
    - **Resource Pooling**: The provider's resources (e.g., storage, processing) are pooled to serve multiple consumers.
    - **Location Independence**: The exact physical location of the resources is irrelevant to the user.
    - **Rapid Elasticity**: Resources can be scaled rapidly, both up and down, to meet demand.
    - **Measured Service**: Resource usage is monitored, controlled, and reported, ensuring users are billed based on consumption.
    - **Multi-Tenancy**: Multiple users share resources without affecting one another.

2. Basic Service Models

- **Software as a Service (SaaS)**:
    - End users access applications running on the cloud.
    - Examples include email services and cloud storage like Google Drive.
    - **Use Case**: An organization uses a cloud-based CRM like Salesforce, enabling its sales team to access customer data from anywhere.
- **Platform as a Service (PaaS)**:
    - Developers deploy applications using programming languages and tools provided by the cloud provider.
    - Examples include Google App Engine and Microsoft Azure App Services.
    - **Use Case**: A developer builds a web app using Microsoft Azure's PaaS offerings, using pre-configured databases, development tools, and APIs.
- **Infrastructure as a Service (IaaS)**:
    - Users can provision processing, storage, and networks to run operating systems and applications.
    - Examples include Amazon EC2 and Microsoft Azure VMs.
    - **Use Case**: A company runs its own custom applications on virtual machines provided by Amazon EC2, with control over OS and installed software.

3. Deployment Models

- **Private Cloud**:
    - Exclusive use by a single organization, ensuring high security and control.

    - **Use Case**: A government agency uses a private cloud for sensitive data management.
- **Public Cloud**:
    - Available to the general public, owned by third-party providers like AWS or Microsoft Azure.
    - **Use Case**: Startups host their websites on public clouds for cost-effectiveness.
- **Community Cloud**:
    - Shared by multiple organizations with similar requirements (e.g., compliance, mission).
    - **Use Case**: Healthcare organizations share a community cloud to manage patient data while complying with healthcare regulations.
- **Hybrid Cloud**:
    - A combination of private, public, or community clouds, ensuring flexibility.
    - **Use Case**: An e-commerce site uses public cloud for its customer-facing services while maintaining a private cloud for sensitive transaction data.

4. Economic Justification

- **Economies of Scale**:
    - Large data centers (100,000+ servers) are more cost-efficient than small ones (<10,000 servers).
    - **Cost Factors**:
        - **Power Costs**: Lower per-server costs due to shared infrastructure, bulk power negotiation, and cheaper power sources like wind or solar energy.
        - **Labor Costs**: More efficient system administration, with one administrator managing over 1,000 servers in large data centers versus 150 in smaller centers.
        - **Security and Reliability**: Larger centers can distribute fixed costs of security, redundancy, and disaster recovery over more servers.
        - **Hardware Costs**: Discounts of up to 30% on hardware for large data center operators.
- **Utilization of Equipment**:
    - Virtualization allows multiple applications to co-locate on the same hardware, improving server utilization.
    - **Factors Enhancing Utilization**:
        - **Random Access**: Random user access results in uniform server load.
        - **Time of Day**: Co-location of workplace and consumer services adjusts workloads by time zones.
        - **Time of Year**: Seasonal demand adjustments (e.g., tax season) help manage resource allocation.
        - **Resource Usage Patterns**: Co-locating CPU-intensive services with I/O-intensive services ensures balanced resource consumption.
        - **Uncertainty**: Cloud services can manage unexpected spikes due to news events or marketing campaigns.

    - **Use Case**: An online streaming service uses a public cloud to manage sudden spikes in demand during popular events like sports finals.
- **Multi-Tenancy**:
    - A single application serves multiple users, reducing costs in:
        - **Help Desk Support**: Centralized support for a single app.
        - **Simultaneous Upgrades**: One update applies to all users.
        - **Unified Development**: Easier maintenance with one version of software.
    - **Use Case**: A SaaS CRM platform hosts a single instance of its application for different businesses, lowering maintenance costs and simplifying updates.

5. Summary

- Cloud computing offers a scalable, flexible, and cost-effective platform for various applications.
- It supports different service models (SaaS, PaaS, IaaS) and deployment models (Private, Public, Community, Hybrid).
- Economic advantages include economies of scale, increased equipment utilization, and multi-tenancy, making it attractive for organizations aiming to optimize costs and efficiency.