

# CLOUD COMPUTING

**Moving from local to remote**

COS216  
AVINASH SINGH  
DEPARTMENT OF COMPUTER SCIENCE  
UNIVERSITY OF PRETORIA



# WHAT IS CLOUD COMPUTING

- On-demand availability of computer system resources and computing power (servers, storage, databases, networking, software, analytics, and intelligence), without direct active management by the user.
  - Over the internet
- Benefits:
  - lower operating costs
  - efficient use of infrastructure
  - disaster recovery
  - economies of scale

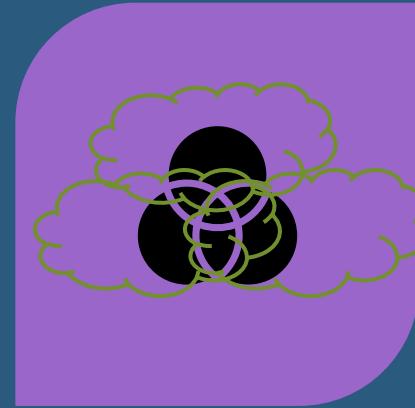
# **TYPES OF CLOUD COMPUTING**



**PUBLIC**



**PRIVATE**



**HYBRID**

# TYPES OF CLOUD COMPUTING

- **Public clouds**
- owned and operated by third-party cloud service providers, which deliver their computing resources over the Internet for public use (likely free).
- Examples:
  - Amazon Web Services (AWS)
  - IBM Cloud
  - Oracle
  - Microsoft
  - Google
  - Alibaba

# **TYPES OF CLOUD COMPUTING CONTD.**

- **Private clouds**
- Computing resources used exclusively by a single business or organization.
- A private cloud is one in which the services and infrastructure are maintained on a private network.
- Can be physically located in the company's on-site data center.
- Some companies also pay third-party service providers to host their private cloud.

# CONTD.

- **Hybrid clouds**
- combine public and private clouds
- bound together by technology that allows data and applications to be shared between them.
- allowing data and applications to move between private and public clouds
- a hybrid cloud gives your business greater flexibility, more deployment options, and helps optimize your existing infrastructure, security, and compliance.

# OTHERS

- **Community cloud** - shares infrastructure between several organizations from a specific community with common concerns (security, compliance, jurisdiction, etc.)
- **Distributed cloud** – creating a cloud platform assembled from a distributed set of machines in different locations, connected to a single network or hub service.
- **Multi-cloud** - the use of multiple cloud computing services in a single heterogeneous architecture to reduce reliance on single vendors.
- **Poly cloud** - the use of multiple public clouds for the purpose of leveraging specific services that each provider offers.
- **HPC cloud** - the use of cloud computing services and infrastructure to execute high-performance computing (HPC) applications. These applications consume considerable amount of computing power and memory and are traditionally executed on clusters of computers

# CLOUD COMPUTING SERVICES

- Most cloud computing services fall into three broad categories: infrastructure as a service (IaaS), platform as a service (PaaS), and software as a service (SaaS).
- Sometimes called the cloud computing "stack" because they build on top of one another
- Knowing what they are and how they're different makes it easier to accomplish your business goals.

# INFRASTRUCTURE AS A SERVICE (IAAS)

- Rent IT infrastructure - servers and virtual machines (VMs), storage, networks, operating systems - from a cloud provider on a pay-as-you-go basis.

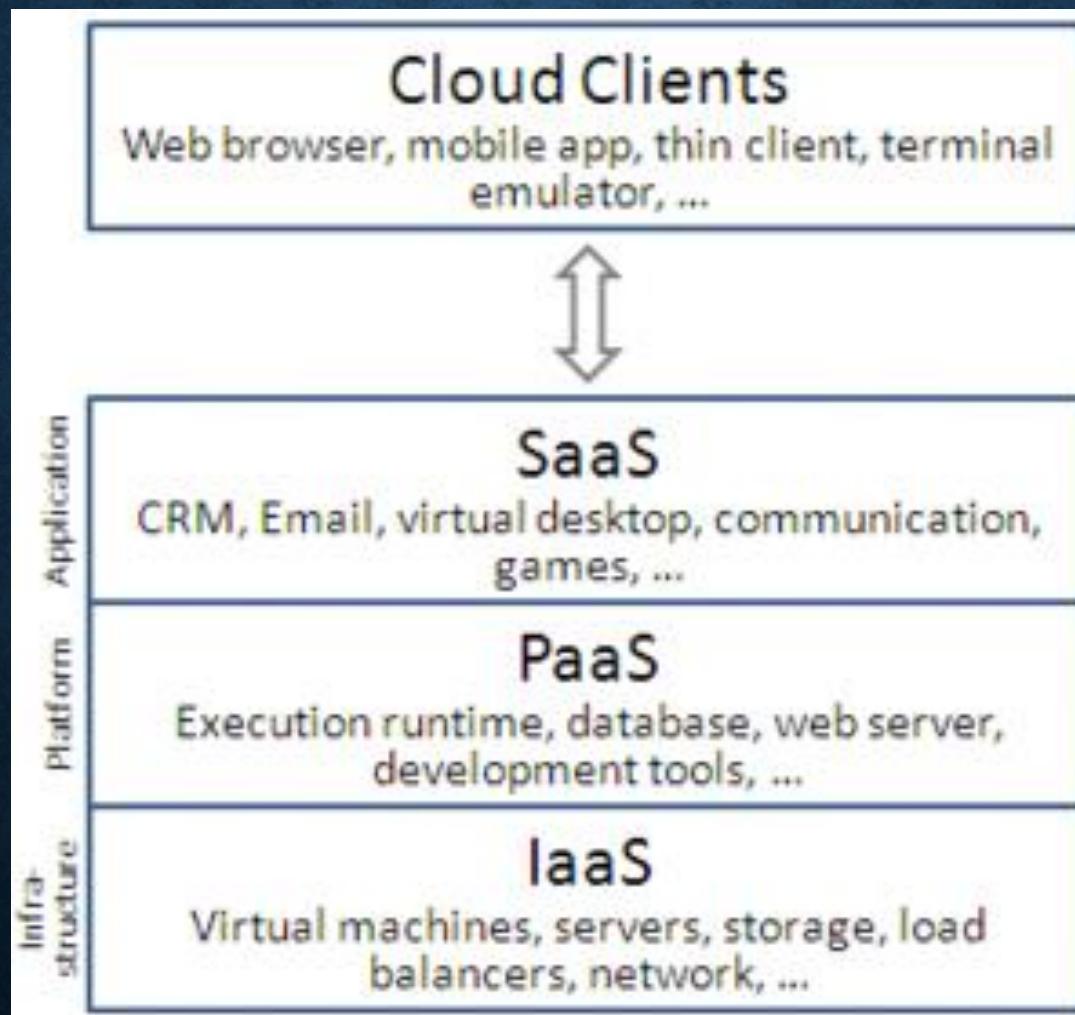
# **PLATFORM AS A SERVICE (PaaS)**

- Services that supply an on-demand environment for developing, testing, delivering, and managing software applications.
- Designed to make it easier for developers to quickly create web or mobile apps, without worrying about setting up or managing the underlying infrastructure of servers, storage, network, and databases needed for development.

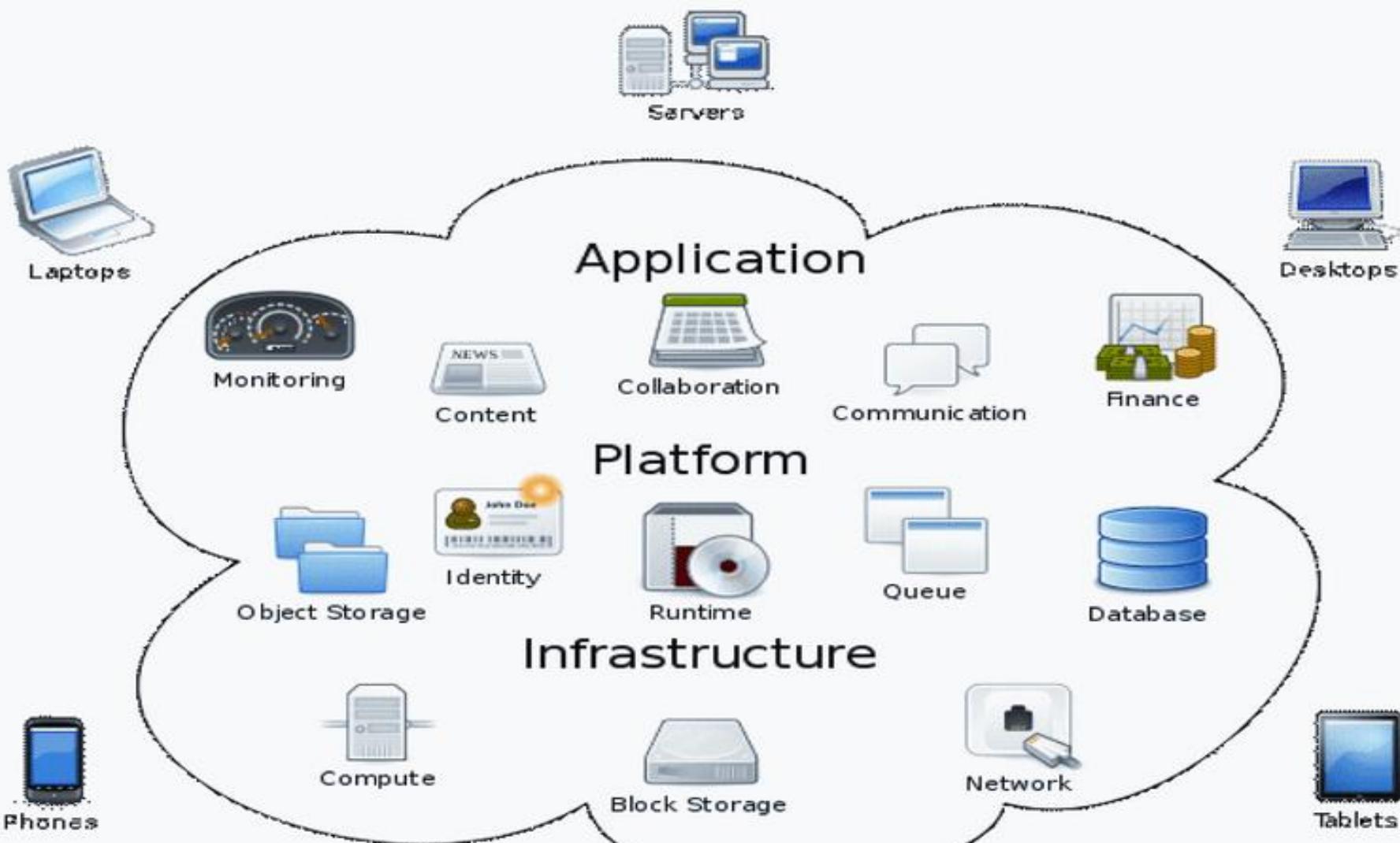
# **SOFTWARE AS A SERVICE (SAAS)**

- A method for delivering software applications over the Internet, on demand and typically on a subscription basis.
- With SaaS, cloud providers host and manage the software application and underlying infrastructure, and handle any maintenance (software upgrades and security patching).
- Users connect to the application over the Internet, usually with a web browser on their phone, tablet, or PC.

# CLOUD COMPUTING STACK



# Cloud Computing



# HOW CLOUD COMPUTING WORKS?

- The cloud system can be divided into two sections: the **front end** and the **back end**.
- They are connected to each other through a network, usually the Internet.
- The front end is the side of the computer user or client.
  - Comprising client devices, networks and applications
- The back end is 'the cloud' section of the system.
  - Comprising computers, servers, data storage, data centers

# SCALING APPLICATIONS IN THE CLOUD

- Scale up (Vertical Scaling) – improving the infrastructure
  - Eg. Upgrading vm from 8GB RAM 4Core -> 512GB RAM 64Core
- Scale out (Horizontal Scaling) – creating multiple instances of the infrastructure
  - Eg. Upgrading vm from 8GB RAM 4Core -> 2 x 8GB RAM 4Core
- Scaling Techniques:
  - Load Balancing/Traffic Management - distribute network traffic load between multiple instances so that no one will fail.
  - Content Delivery Network (CDN) - to reduce the latency for getting static content from the server to the user's location.

# USING CLOUD COMPUTING

- Internet connection and bandwidth
  - Its implication for clients
- Hardware consideration
- Service Level Agreements (SLA)
  - Demand clear and detailed information from provider
- Disaster Recovery
  - What are the policies and strategies for recovering

# USING CLOUD COMPUTING

- Data Location
  - Laws and regulations of the jurisdictions where data may reside
- Scalability
- Sustainability
- Security
- Cloud Service Management

# DATA IN THE CLOUD

- Cloud storage is a model of computer data storage in which digital data is stored in logical pools.
- The physical storage spans multiple servers (sometimes in multiple locations), and the physical environment is typically owned and managed by a hosting company.
- These cloud storage providers are responsible for keeping the data available and accessible, and the physical environment protected and running.

# DATA IN THE CLOUD

- People and organizations buy or lease storage capacity from the providers to store user, organization, or application data.
- Cloud storage services may be accessed through a colocated cloud computing service, a web service application programming interface (API) or by applications that utilize the API, such as cloud desktop storage, a cloud storage gateway or Web-based content management systems.

# BEFORE SAVING IN THE CLOUD

- Data security
  - Replication across various location
- Data longevity
- Data accessibility
  - Performance trade-off between remote vs. local storage
  - Availability and reliability delegated

