Cloud Computing

Introduction

- The seed- Network communication
- Germination-Internet
- Growth- Distributed/grid computing
- Reproduction- cluster computing ,
- Pollination-cloud computing
- Spreading of seeds –fog and edge computing

Dominance of cloud and data center demands

- Arm eco system
- Web applications, , social networking, cloud apps
- Machine learning
- Big data analytics(big data processing engines apache spark or presto)
- HPC
- Speech recognition, real time video streaming, 3 D visualization and graphic intensive workloads
- Data warehousing
- Parallel processing,
- ERP, **Internet of Things (IoT)**
- Map reduce and Hadoop distributed computing

Characteristics

- Large scale Shared IT infrastructure
- On demand self service
- Large network access
- Availability (scalability, elasticity, and dynamic scheduling)
- Pay as you go
- Customization
- Reliability (Service Level Agreement)
- Efficient resource utilization
- Energy efficient (go Green)
- Automated and Well managed Virtual environment
- Any Application workloads
- Geographic Distribution
- No Transparency and end to end control

Advantages of cloud computing

- *Scalability*
  - *Storage*
  - *computing*
- *economical*
- *Simplicity*
  - *easy to access,*
  - *Quick Deployment with compatible infrastructure*
  - *Automatic software integration ,updation*
  - *Back up and recovery*

Disadvantages

- Internet Access
- Security and privacy challenges
- ***Misconfigured servers are responsible for 86% of compromised records.***
- ***API vulnerability***
- ***Malicious insider activity***

CAM4—2020

The leaked database included location details, email addresses, IP addresses, payment logs, usernames and more. Exposed 10.8 billion sensitive entries amounting to 7 TB of data.

Advanced Info Service (AIS)—2020

Included 8.3 billion network flow logs and DNS query logs

Keepnet Labs—2020

The database included two data collections containing 5 billion and 15 million entries exposed to various breaches

Microsoft—2019

The exposure of 250 million entries, including email addresses, IP addresses, and support case details.

(Reference : https://www.triskelelabs.com/blog/cloud-cyber-attacks-the-latest-cloud-computing-security-issues

- Lack of standards

  interoperability is difficult (to move workloads between private and public cloud)
- Compliance concerns
  - Lack of full control and transparency
  - Protection laws differ from one country to another
- Service migration
  - No external interface
  - Interaction between cloud providers
- Evolving nature

System Architectures

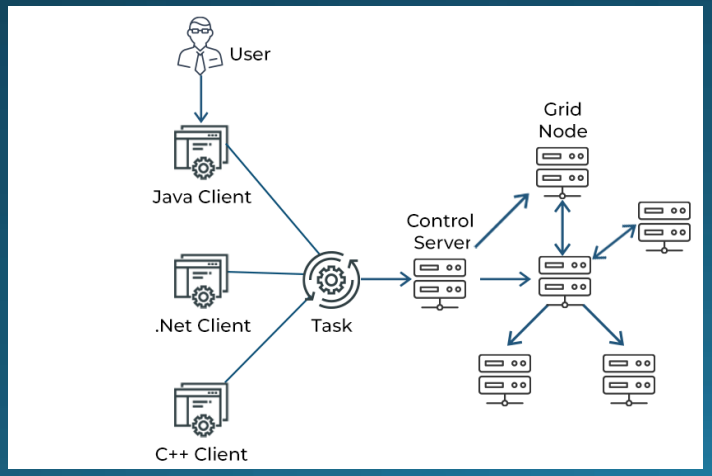
- Centralized systems
- Decentralized systems
- Distributed systems
  - Properties:
    - Load sharing
    - Each node play a partial role
    - Resource sharing
    - Load sharing
    - Easy to expand
    - Performance(parallel computing- compute and data intensive App)
    - No single point failure

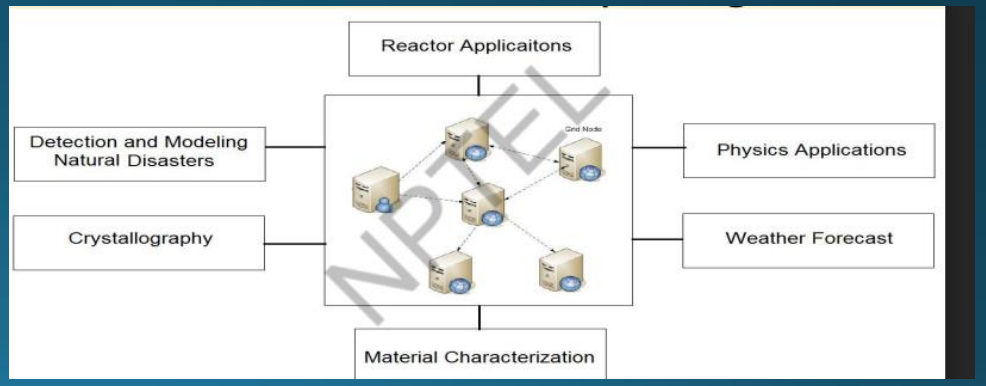
Key pointe to be compared:
Maintenance
Security
Scalability and availability
Cost Effectiveness
Fault tolerant and reliable
Traffic congestion
Fast and easy deployment
Performance
Load balancing

Grid Computing

- Share more than information: Data, computing power, applications in dynamic environment, multi-institutional, virtual organizations
- Efficient use of resources at many institutes. People from many institutions working to solve a common problem (virtual organisation).
- Join local communities.

Interactions with the underneath layers must be transparent and seamless to the user.





- Grid **computing**
  - **Tightly coupled**
  - **Distributed scheduling and job management**
  - **No SSI**

Advantages

Availability scalability speed

Disadvantages

Single point failure

very complex

- **Cluster computing**
    - **Tightly coupled**
    - **Centralized job management and scheduling**
    - **SSI**

advantages

- **Availability scalability speed**
- **Scalability**
- **High Performance**
- **eliminates single point failure**

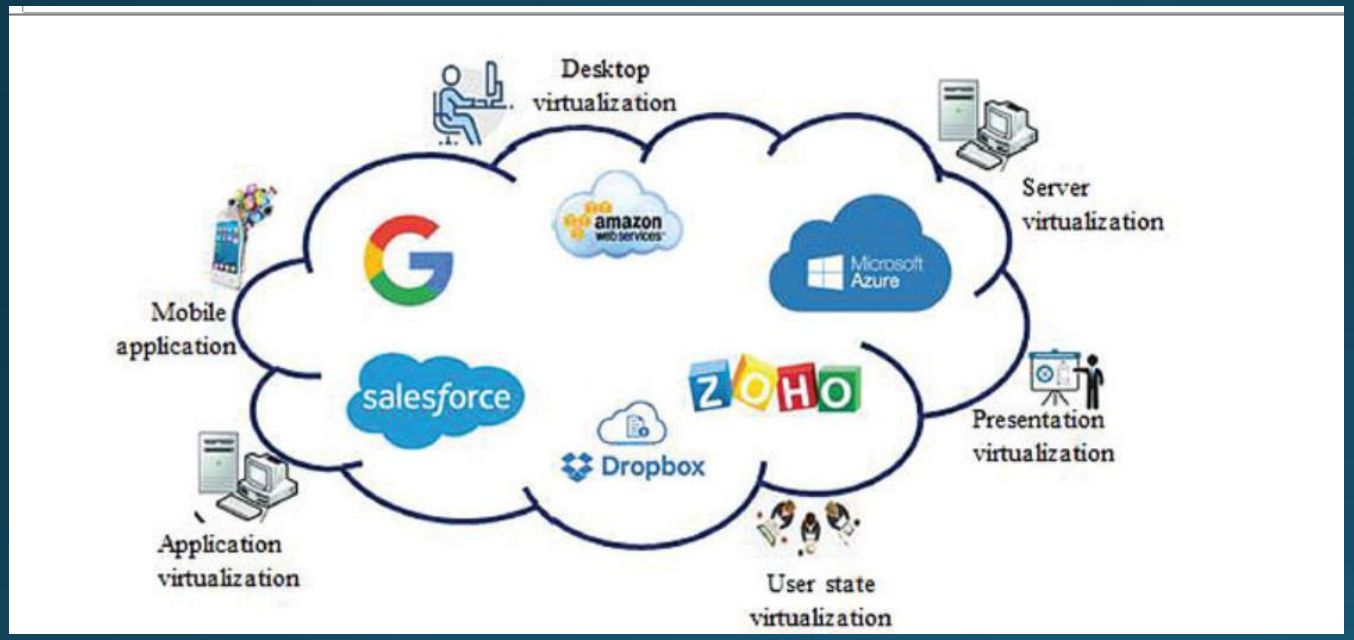
Disadvantage:

- **finding fault is difficult**

Examples of cluster computing

- Aerodynamics and data mining applications use the clustering process.
- Fast image rendering can also be done by using the clustering technique.
- It is also used in weather forecasting techniques.

Cloud computing



Traditional and cloud computing

Characteristics	Clusters	Grids	Clouds
Ownership	Single ownership	Multiple ownership	Single ownership
Service pricing	Limited	Private or public assigned	Utility /large user discount
Virtualization	Half	Half	Yes
Resource management	Centralized resource	Distributed resource	Both
Scalable size	100s	1000s	100 to 1000s
Standardized	Yes	Yes	No
Interoperability	Yes	Yes	Not full
Speed / Interconnected network	Dedicated high end with low latency and high bandwidth	Mostly internet with high latency and low bandwidth	Dedicated high end with low latency and high bandwidth

Self-service	No	Yes	Yes
Single system image	Yes	No	Yes/optional included
Multi-tenancy	No	Yes	Yes
Service negotiation	Limited	Yes, SLA-based	Yes, SLA-based
Membership discovery	Membership service discovery	Decentralized information services and centralized indexing	Membership service discovery
Operating system	Windows/Linux	Any standard but dominated by Unix	Uses a hypervisor
Application drivers	Business, data centres, enterprise computing	Collaborative scientific and high-throughput applications	Web App. content delivery, dynamic provisioning

Capacity	guaranteed capacity	Varies, but high capacity	Provisioned on-demand capacity
Security	Traditional login/ password-based	Public/private pair -based authentication and mapping of a user to an account	Each user and/ or application is provided with a virtual machine
Privacy	Medium level of privacy depends on user privileges	Limited support for privacy	High security/ privacy is guaranteed. There is support for file **Access Control List** (**ACL**) settings.
Population	Commodity computers	High-end computing systems (including clusters and servers)	Commodity PCs, high-end servers' network, attached storage

Future of Cloud

- Enabling next generation data centers
- Managing Hybrid Cloud
- Every thing as a Service

Cloud Deployment Models

- Public/External Cloud
  - Characteristics of cloud
  - Advantages of cloud
  - Disadvantages of cloud
  - User: Enterprise/ Organization / Public user
- Private/internal/corporate cloud
  - On site

    Up front cost: Cloud management software installation.
  - Off site(out sourced/hosted)
    - the resources are provisioned by the Provider
    - Upfront cost:
    - Negotiating the terms of the service level agreement (SLA)
    - Possibly upgrading the subscriber's network to connect to the outsourced private cloud
    - Switching from traditional applications to cloud-hosted applications,
    - Porting existing non-cloud operations to the cloud
    - Training

Explore all the cloud deployment models based on characteristics In terms of

Scalability, elasticity, availability
Storage Size
Accessibility
Infrastructure
Ownership
Workloads
Well managed virtual environment
Customizable
Reliability
Cost effectiveness

- Advantages of Private cloud
  - Security from outside threats and privacy
  - Full control
  - Low latency
  - Predicted cost
  - Reliable
  - Direct Access, Independent
- Disadvantages
  - Limited scalability
  - Specific application workload
  - Initial investment
  - Skilled IT Professional (Maintenance)
  - Performance limitation with limited resources

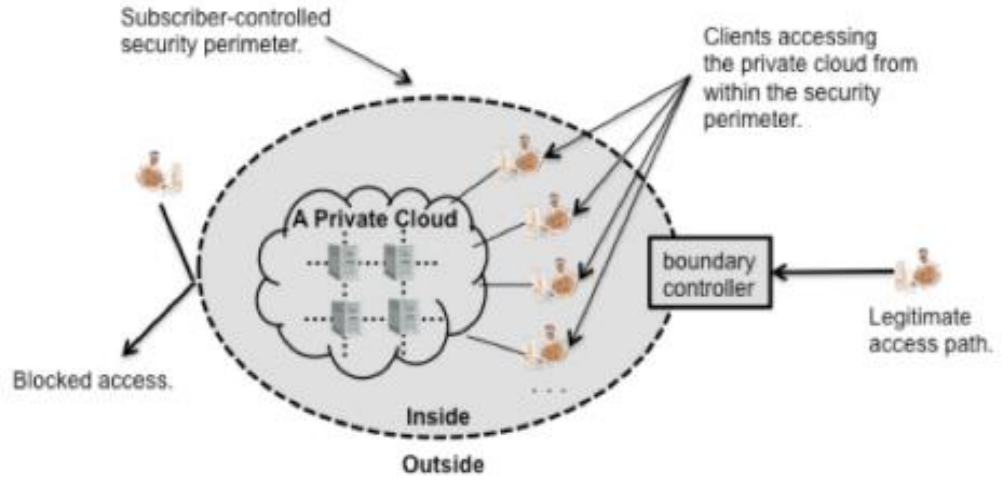


Figure 3: On-site Private Cloud

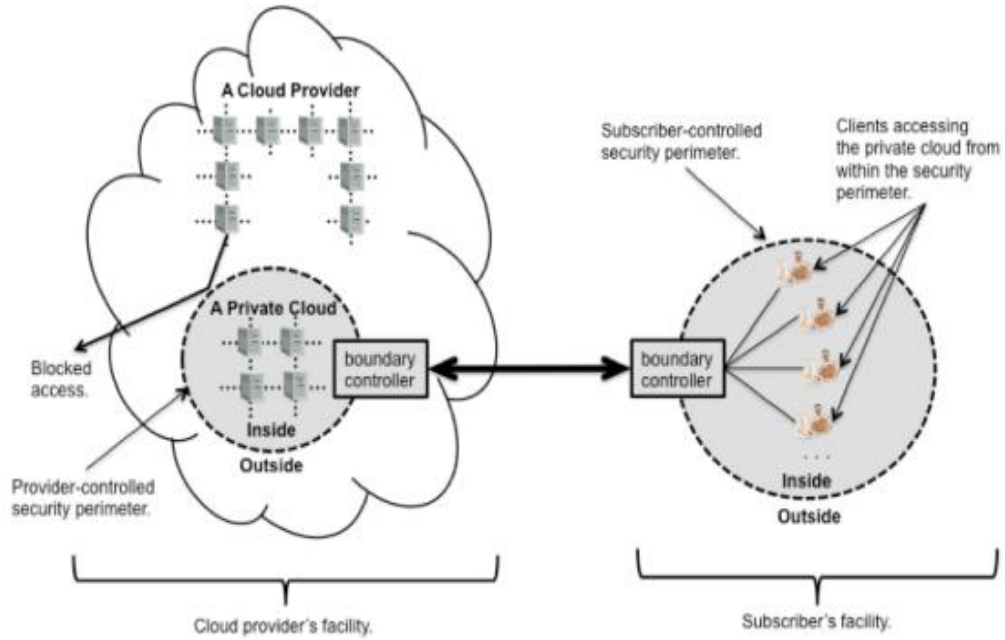


Figure 4: Outsourced Private Cloud

- Hybrid cloud
  - To maximize disaster recovery speed
  - Data heavy work loads
    - **Heterogeneous**
    - Homogeneous

      Easy to use often means harder to leave
    - Ensures application and data portability
    - Private (critical services)
    - Public (back up – cost effective scalability)
    - Data Application portability standardization

  Advantages:

    - Unique balance of control, performa
    - Business agility
    - Cost savings
    - Fast deployment
  - **Disadvantages**
    - **Complex security**
    - **Expertise fro successful integration**
  - Easy hybrid cloud deployment
  - Better Performance
- Inter-Cloud

Ref:https://csrc.nist.gov/csrc/media/publications/sp/800-1

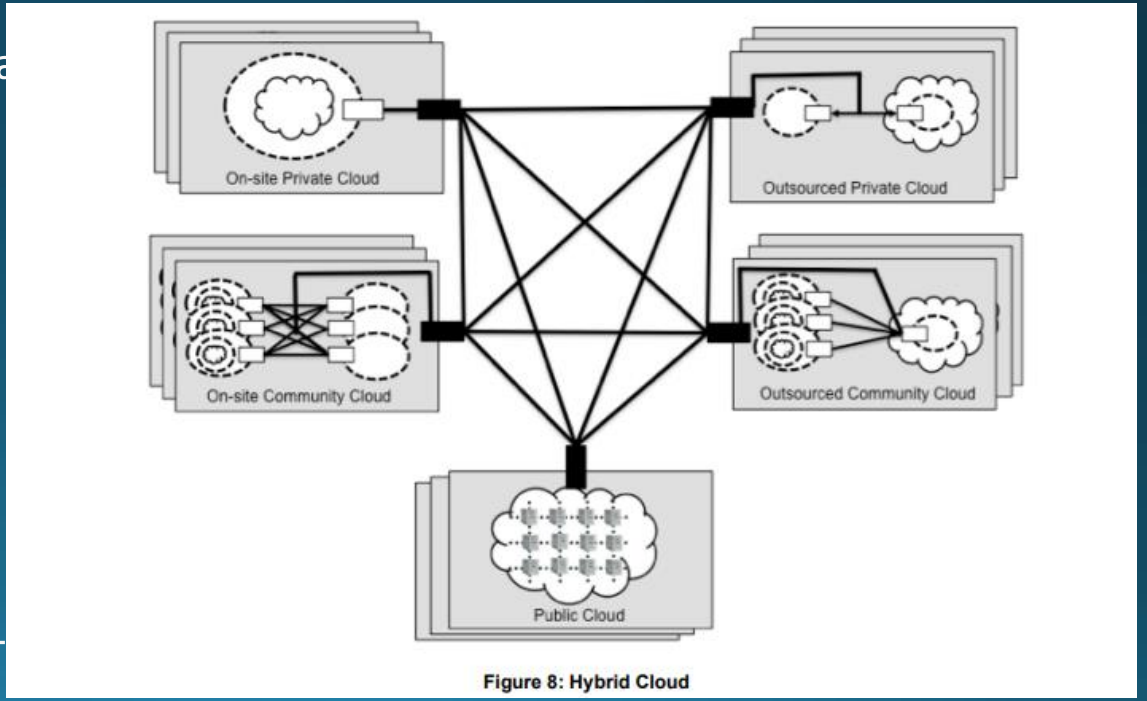


Figure 8: Hybrid Cloud

- Community cloud
  - Private
  - Public

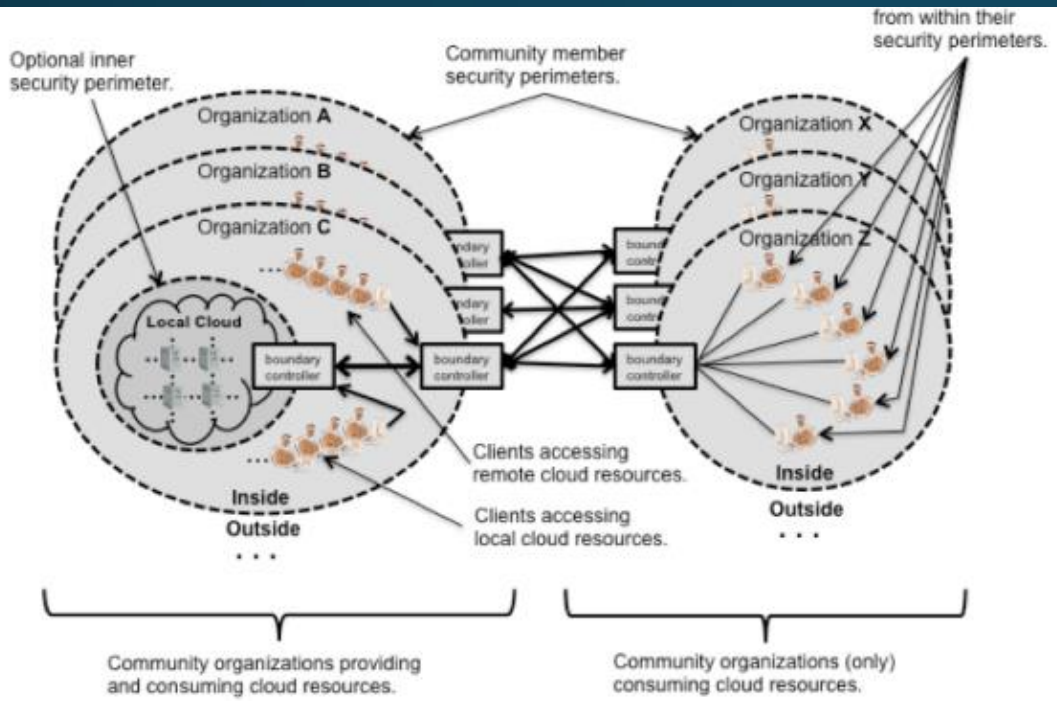
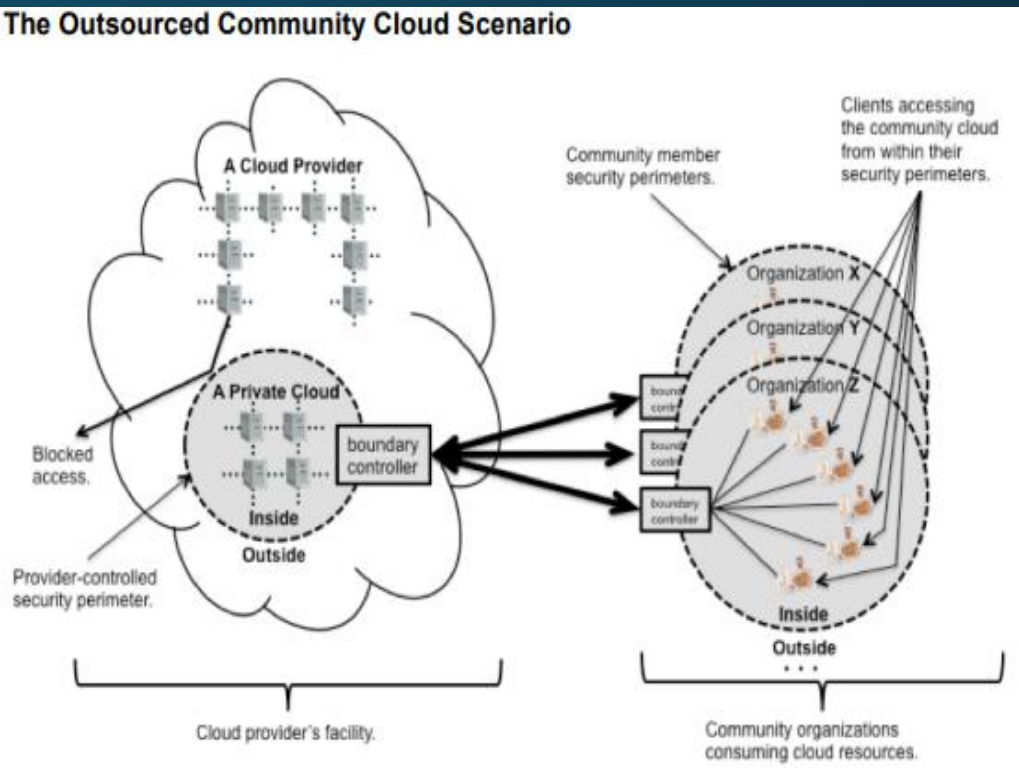


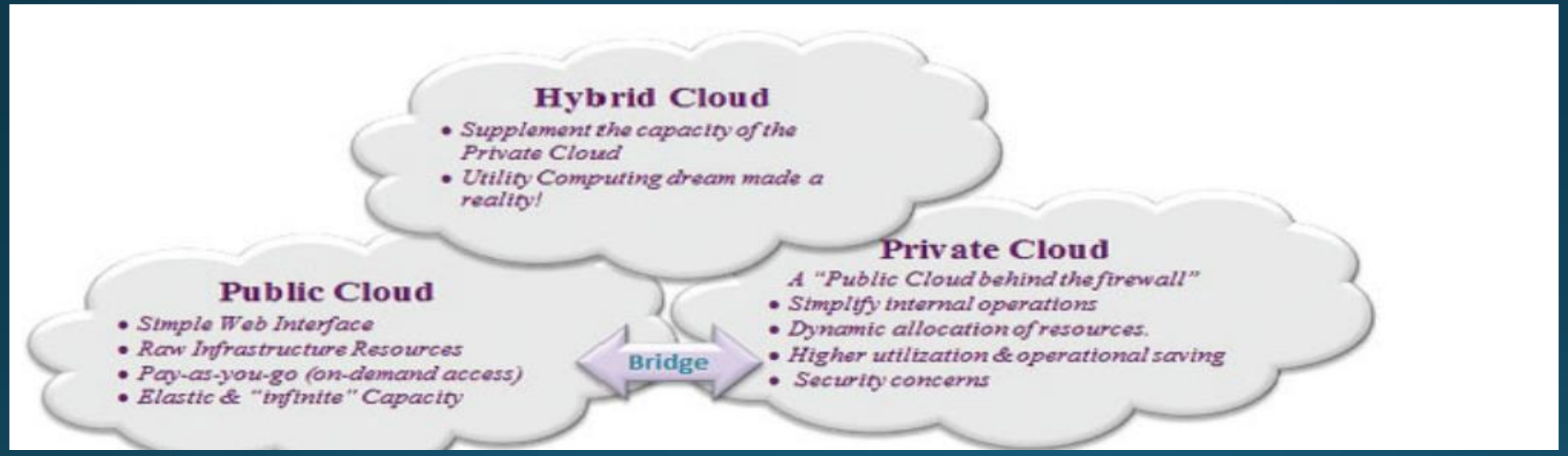
Figure 5: On-site Community Cloud



- Public Cloud:-Rackspace, **Amazon Elastic,Compute Cloud (Amazon EC2), Sun Cloud, IBM's Blue Cloud,** Windows Azure Services Platform
- Private cloud:**Simple Storage Service,(S3), Red Hat, SaM CloudBOX and Amazon EC2.**

- **Private cloud providers:**
  - **Hewlett Packard Enterprise**
  - **Cisco**
  - **Microsoft**
  - Other major private cloud providers include Dell, IBM, VMware, Oracle, and Red Hat.

    Each provider has a unique set of cloud offerings.
- Community cloud: Google Government Cloud, IGT Cloud, Optum Health Care Cloud,
- Hybrid Cloud:
  - Windows Azure (capable of Hybrid Cloud)
  - – VMware vCloud (Hybrid Cloud Services)


Hybrid Cloud
• Supplement the capacity of the Private Cloud
• Utility Computing dream made a reality!
Public Cloud
• Simple Web Interface
• Raw Infrastructure Resources
• Pay-as-you-go (on-demand access)
• Elastic & "infinite" Capacity
Bridge
Private Cloud
A "Public Cloud behind the firewall"
• Simplify internal operations
• Dynamic allocation of resources.
• Higher utilization & operational saving
• Security concerns

Cloud Service Models

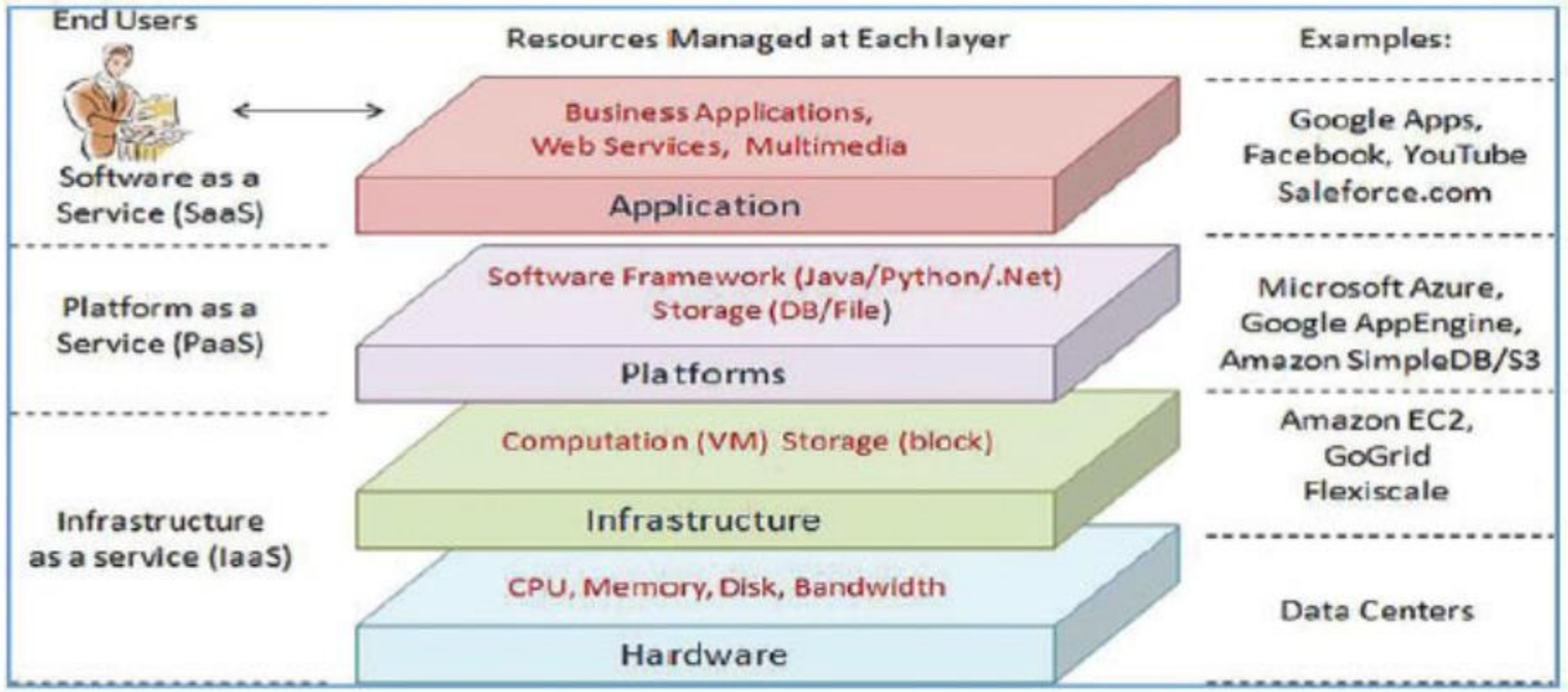
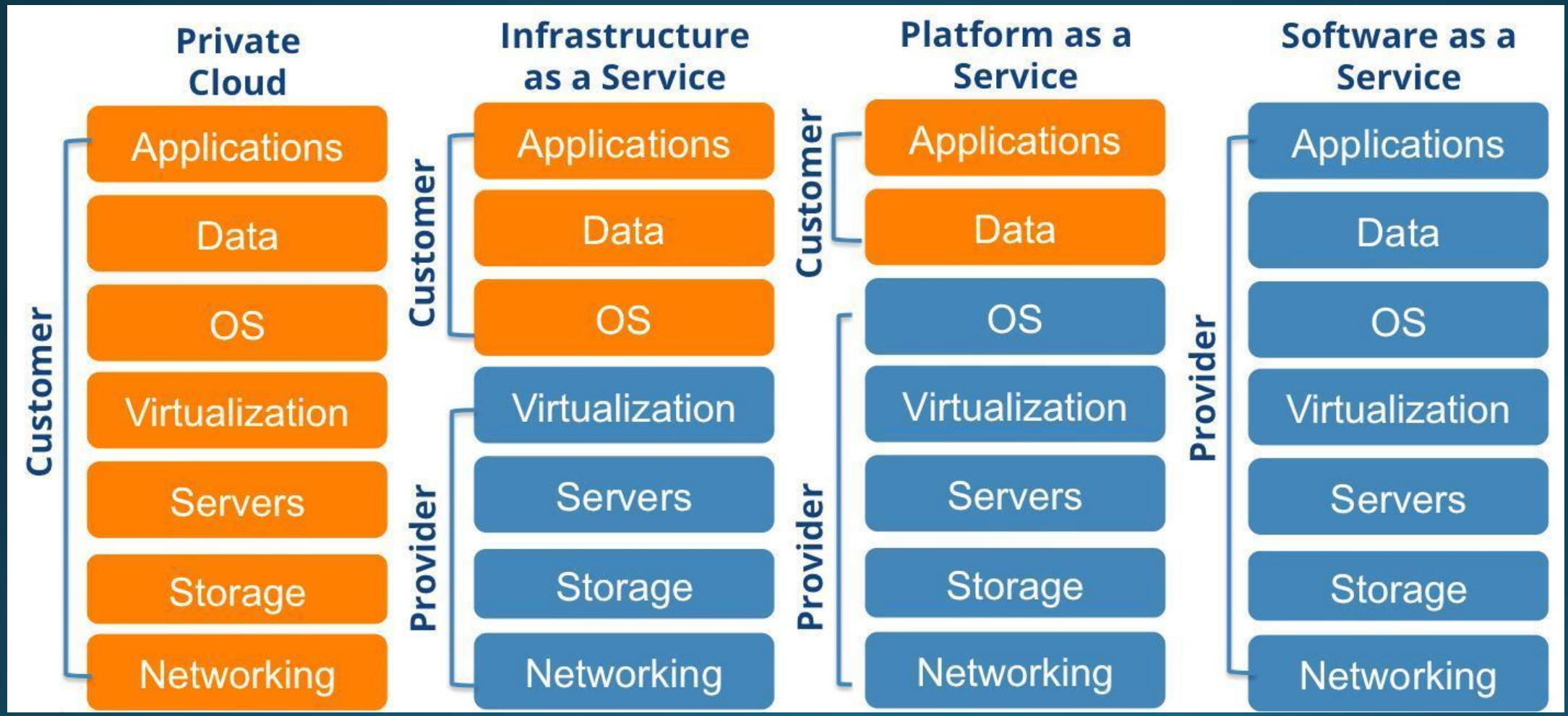


Figure A layered model of cloud services


Private Cloud
Customer
Applications
Data
OS
Virtualization
Servers
Storage
Networking
Infrastructure as a Service
Customer
Applications
Data
OS
Provider
Virtualization
Servers
Storage
Networking
Platform as a Service
Customer
Applications
Data
Provider
OS
Virtualization
Servers
Storage
Networking
Software as a Service
Provider
Applications
Data
OS
Virtualization
Servers
Storage
Networking

IaaS

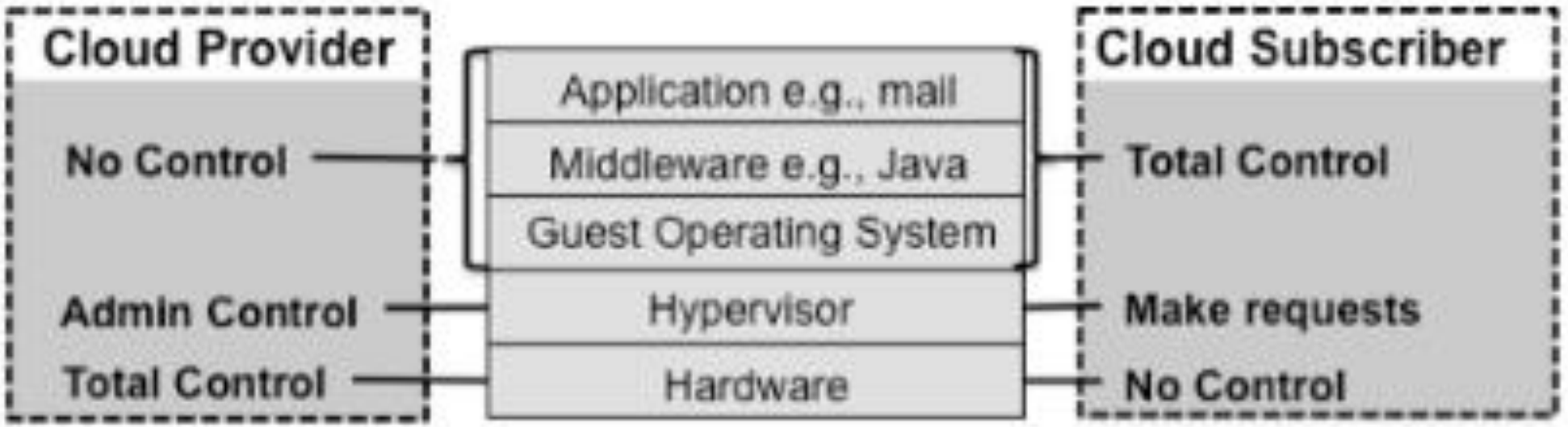
- Virtual private server
- Service Provider API

Characteristics

- Resources are provisioned as a service
- Availability is based on a utility-based pricing model
- Scaling of resources is dynamic and on-demand
- Supports multiple users to concurrently access a single piece of hardware

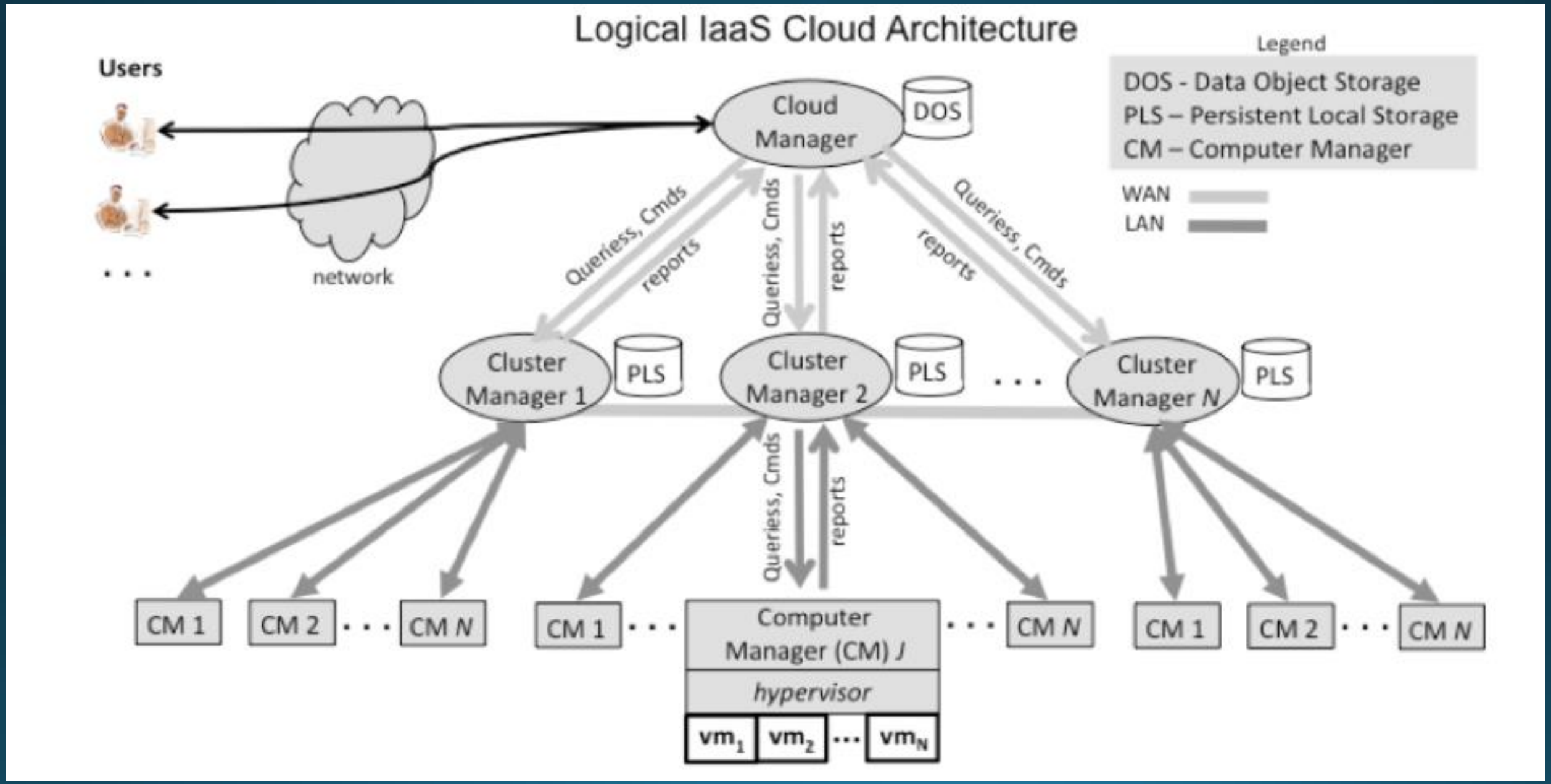
IaaS Component Stack and Scope of Control

The primary units of allocation are (administrative access to) VMs, network bandwidth, storage, and IP addresses. Additional resources include monitoring services, firewalls, synchronization mechanisms such as queues, databases, etc.



Three-level hierarchy of components in IaaS cloud systems

- – *Top level is responsible for central control*
- – *Middle level is responsible for management of possibly large computer clusters* ***that may be geographically distant from one*** another
- – *Bottom level is responsible for running the host computer systems on* which virtual machines are created.


Logical IaaS Cloud Architecture
Users
network
Cloud Manager
DOS
Queriess, Cmds
reports
Queriess, Cmds
reports
Queriess, Cmds
reports
Legend
DOS - Data Object Storage
PLS – Persistent Local Storage
CM – Computer Manager
WAN
LAN
Cluster Manager 1
PLS
Cluster Manager 2
PLS
Cluster Manager N
PLS
Queriess, Cmds
reports
CM 1
CM 2
CM N
CM 1
Computer Manager (CM) J
hypervisor
vm1
vm2
vmN
CM N
CM 1
CM 2
CM N

Cloud Manager

- Cloud Manager is the public access point to the cloud where subscribers sign up for accounts, manage the resources they rent from the cloud, and access data stored in the cloud.
- Cloud Manager has mechanism for:
  - Authenticating subscribers
  - Generating or validating access credentials that subscriber uses when communicating with VMs.
  - Top-level resource management.
- For a subscriber's request cloud manager determines if the cloud has enough free resources to satisfy the request

Data Object Storage (DOS)

DOS generally stores the subscriber's metadata like user credentials, DOS service is (usually) single for a cloud.

Persistent local storage

- Each Cluster Manager is connected to Persistent Local Storage

(PLS)

- PLS provide persistent disk-like storage to Virtual Machine

- **Operation of the Cluster Managers**

- Each *Cluster Manager is responsible for the operation of a collection of* computers that are connected via high speed local area networks
- *Cluster Manager receives resource allocation commands and queries* from the *Cloud Manager, and calculates whether part or all of a* command can be satisfied using the resources of the computers in the cluster.
- *Cluster Manager queries the Computer Managers for the computers in* the cluster to determine resource availability, and returns messages to the *Cloud Manager*
- Directed by the Cloud Manager, a Cluster Manager then instructs the Computer Managers to perform resource allocation, and reconfigures the virtual network infrastructure to give the subscriber uniform access.
- Computer Manger maintains status information including how many virtual machines are running and how many can still be started
- Computer Manager uses the command interface of its hypervisor to start,stop, suspend, and reconfigure virtual machines

Advantages:

Savings in UP-front Cost

- **Full Control of the Computing Resource Through Administrative Access to VMs**
- **Flexible, Efficient Renting of Computing Hardware**
- **Portability, Interoperability with Legacy Applications**

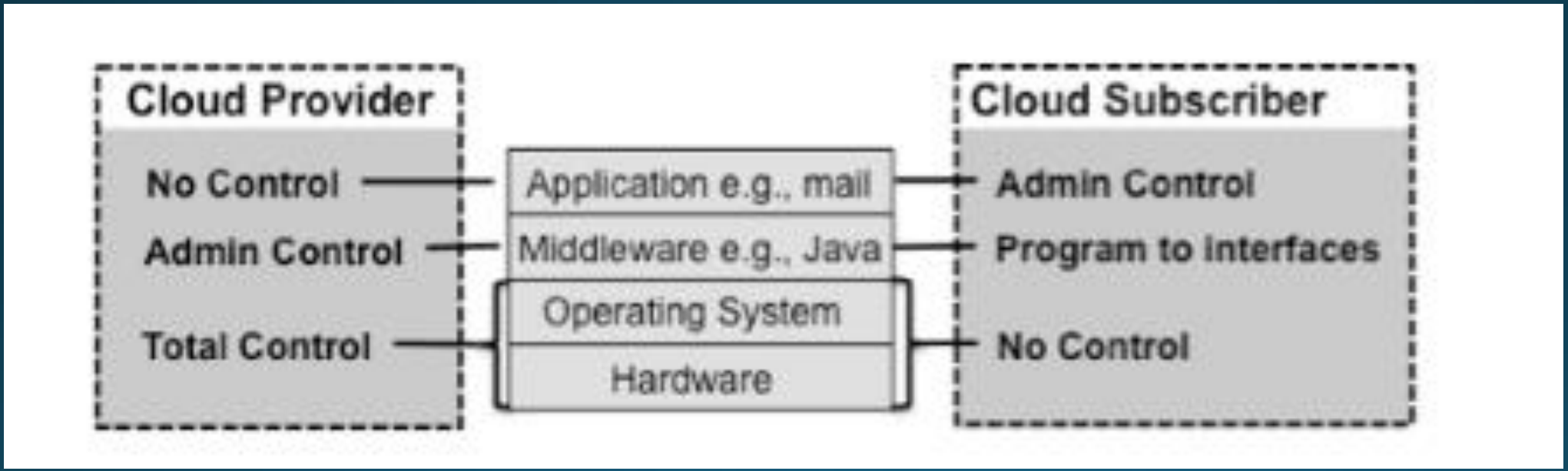
Dis advantages

- **Compatibility with Legacy Security Vulnerabilities**
- **Virtual Machine Sprawl**
- **Verifying Authenticity of IaaS Cloud Provider Web site (web spoofing)**
- **Robustness of VM-level Isolation**
- **Data Erase Practice**

Platform as a Service

- PaaS clouds are similar to any traditional computing system (i.e., platform) in that software applications can be developed for them and run on them.
- Provides a toolkit for conveniently developing, deploying, and administering application software that is structured to support large numbers of subscribers, process very large quantities of data, and potentially be accessed from any point in the Internet.
- Development tools such as programming languages and supporting run-time environments(compilers, interfaces, testing tools, and mechanisms to deploy an application once it's finished). that facilitate the construction of high-quality, scalable applications and deployment .
- Once a subscriber has used the facilities of the PaaS cloud to implement and deploy an application, the application essentially is a SaaS deployment

PaaS component Stack and scope of Control



- Advantages

  Centralized Management and Data

  Platform Issues Managed by Providers

  Savings in Up-front Costs

  Facilitated Scalable Application Development and Deployment

Disadvantages

- Lack of Portability between PaaS Clouds
- Event-based Processor Scheduling
- Security Engineering of PaaS Applications

Software as a Service

- "Software deployed as a hosted service and accessed over the Internet"

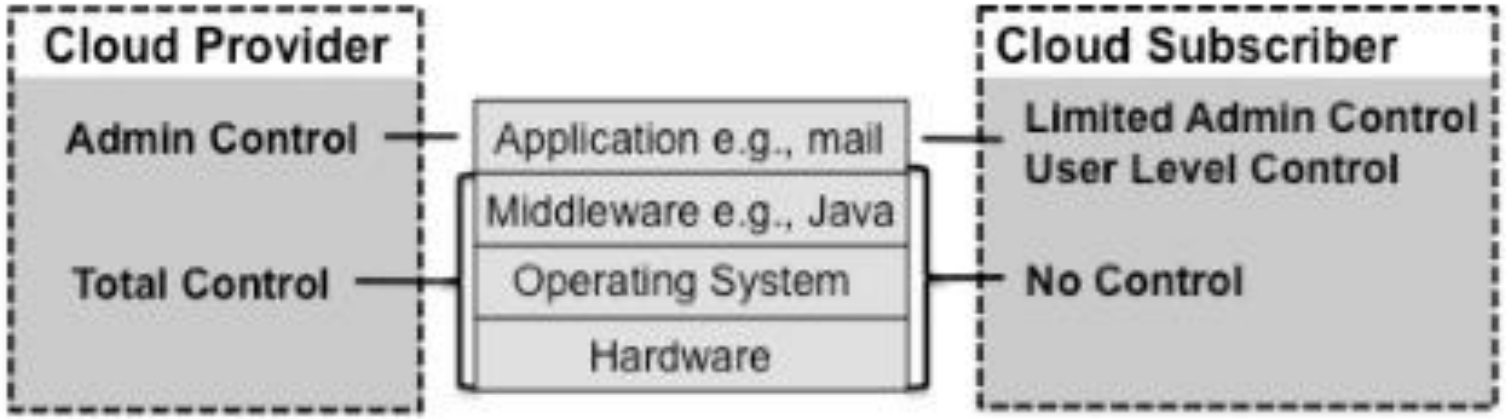
SaaS

Who are the subscribers?

1. Organizations providing their members or employees with access to typical software applications such as office productivity or email.
2. End users who directly use software applications, whether on their own behalf or that of their organization.
3. Software application administrators who configure an application for end users.

What does a subscriber get? The right to use specific applications on demand, and application data management, such as backup and data sharing between subscribers.

How are usage fees calculated? Typically, based on the number of users, the time in use, per-execution, per-record-processed, network bandwidth consumed, and quantity/duration of data stored.



The provider is responsible for deploying, configuring, updating, and managing the operation of the application so that it provides expected service levels to subscribers.

Advantages:

Very Modest Software Tool Footprint

Efficient Use of Software Licenses

Centralized Management and Data

Platform Responsibilities Managed by Providers

Savings in Up-front Costs

Disadvantages:

- **Browser-based Risks and Risk Remediation**
- **Network Dependence**
- **Isolation vs. Efficiency (Security vs. Cost Tradeoffs)**

- Candidate application SaaS and PaaS:
  - Inventory management, Email, Screen sharing, Document sharing, Online gaming, word processor, spread sheet, presentations, format conversion tools, web development..

DATA

- Acquisition
- Preservation
- Organization
- Analysis
- Exchange

Managing Data in cloud (Storage as a Service)

- Basic storage models
  - File system- files and directories
  - BLOB- binary large object- flat object model-extremely scalable
  - Data base-SQL, NoSQL ,graphs
  - Data warehouse
  - Archival storage

Right storage Model

Main factors to be considered

- Nature and size of data
- What type of analysis
- Sharing plan
- Update frequencies

(Performance speed, cost, scalability and reliability characteristics)

File system

- Tree /hierarchical structure
  - POSIX API
    - Apple, Linux, Windows
  - Interface
    - GUI, command line
    - read, write, create, delete
  - Advantages:
  - Direct usage without modification
  - Concurrent access to multiple users
  - Navigate with familiar file system
  - Disadvantage:
  - Relation ship among data is not preserved
  - Experimental results –CSV inconsistent data is not prevented
  - Not suitable for Complex data collections
  - Not scalable
  - Luster- parallel file system by linux cluster community

Object Storage(Blobs)

- Eliminates hierarchy
- Forbids updation
  - Stored and managed as blocks and volumes
  - Self contained repository with metadata
  - Unstructured data
  - Backup and archival

objects are placed in Containers

Objects- unique identifier with metadata

Authorized individual with object identifier can access the data using HTTP request/API

Galcier service-data archival

Examples :- storing objects like videos and photos on Facebook, songs on Spotify, or files in online collaboration services, such as Dropbox.

Advantages:

No updation-Performance improves no synchronization (only read)
Easy to build
Scalable
Managing unstructured data
Disaster Recovery
Less complex

Disadvantages:

Does not support organizing and search- large number of object identifier
Does not provide mechanism to operate with structure data
It cannot be easily mounted like file system

- Relational Database
  - Structured collection and categorization
  - Transactional processing
    - Supports query language
      - Simplifies data management and manipulation
      - Cloud SQL, Cloud Spanner and AlloyDB, amazon Aurora service
      - ACID properties

NoSQL data base

- Not only SQL
  - Unstructured data
  - Does not support joint queries
  - Key-value store-data no represented in tabular form
  - Permits rapid injest of large quantities of unstructured data
  - Distributed over multiple servers and replicated over data centers
  - Eventual consistency
  - Right combination of availability and consistency will depend on the need of the service
  - dynamo DB
  - Cloud bigtable, googles cloud datastore
  - Azure table service
  - Document DB, azure HDInsight

- Graph Database
  - Edges and nodes

  Application

  - Fraud Detection.
  - Recommendation Engines.
  - Network/Operations Mapping.
  - AI Knowledge Graphs.
  - Social Networks.
  - Supply Chain Mapping.
- Data Warehouse
  - Analytical processing
    - Amazon redshift
    - Big query
    - Azure data lake

Table 2.1: Storage as a service options from major public cloud vendors.

Model	Amazon	Google	Azure
Files	Elastic File System (EFS), Elastic Block Store (EBS)	Google Cloud attached file system	Azure File Storage
Objects	Simple Storage Service (S3)	Cloud Storage	Blob Storage Service
Relational	Relational Data Service (RDS), Aurora	Cloud SQL, Spanner	Azure SQL
NoSQL	DynamoDB, HBase	Cloud Datastore, Bigtable	Azure Tables, HBase
Graph	Titan	Cayley	Graph Engine
Warehouse analytics	Redshift	BigQuery	Data Lake

Service Access Methods

Two Access Methods

- Web Portals
  - Not suitable for repetitive Tasks
  - good for simple actions
- RESTAPIs and SDK

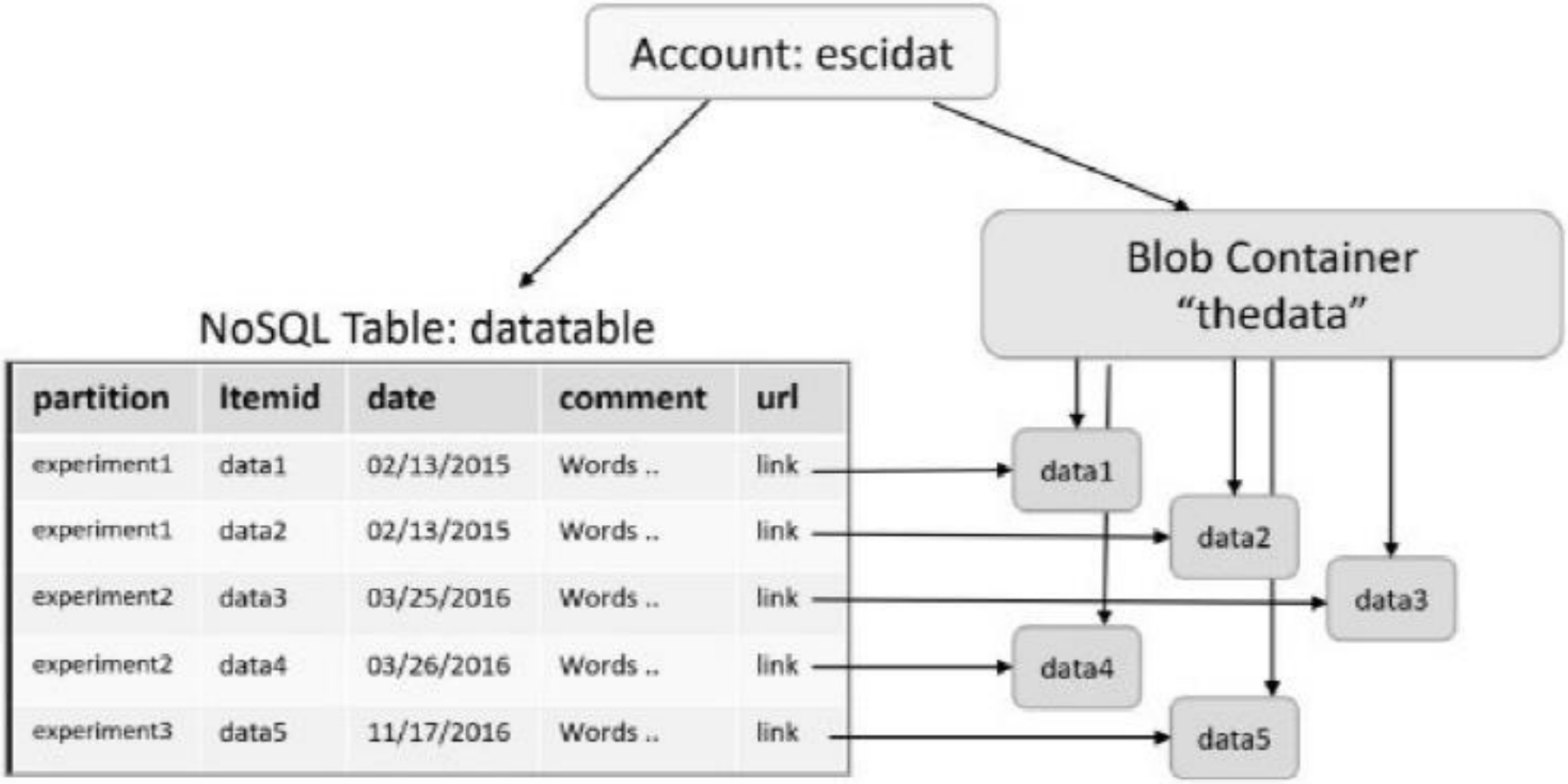


Figure 3.1: The simple cloud storage use case that we employ in this chapter involves the upload of a collection of data blobs to cloud storage and the creation of a NoSQL table containing metadata.

Explore the cloud storage use case (previous slide) using

- Amazon Cloud Service
- Google Cloud Service
- Microsoft Azure