Introduction to Cloud Computing

By

Dr. Shridhar G. Domanal

GTS Labs, IBM, Bengaluru, India

Give a Try

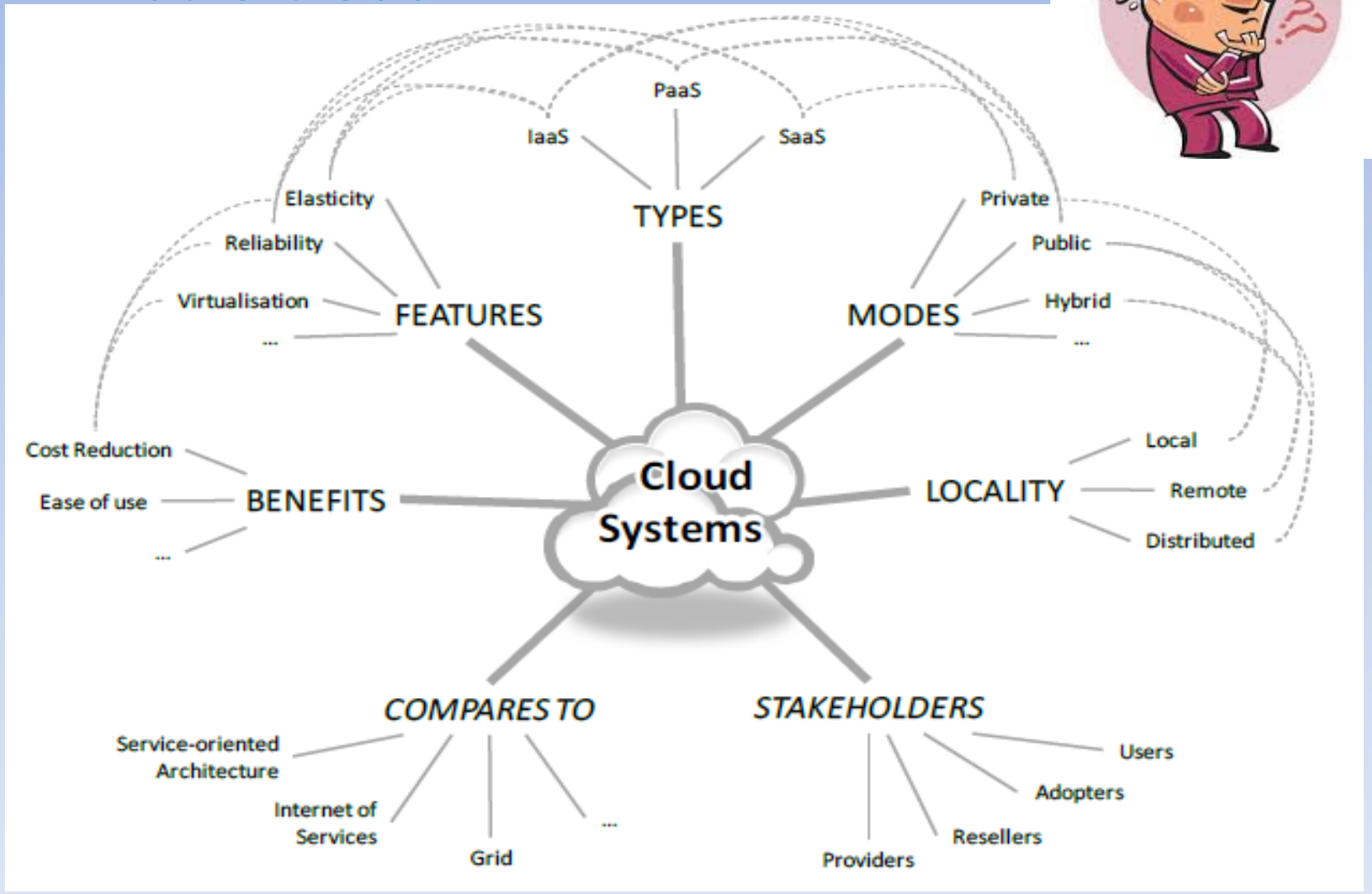
- Network
- Computing
  - Centralized, Distributed, Cluster, Peer-to-Peer, Grid, Cloud, Edge, Serveless, etc.
- Tightly Coupled vs Loosely Coupled
- Protocol
- Image of an Operating System
- Configuration of a Node

AGENDA

- Introduction
- Definition of Cloud
- Typical File Fetching Procedure
- Types of Cloud
- Cloud Architecture
- Cloud Service Models
- Cloud Characteristics
- Role of VM and Hypervisor
- Terminologies in Cloud
- Cloud Simulators
- Issues in Cloud Computing

CLOUD

- What is Cloud?



Introduction

- Network
- Computing
- Centralized Computing
- Distributed Computing
- Cluster Computing
- Grid Computing
- Cloud Computing and many more
- Traditional Vs Cloud (mainly examples)

Definition of Cloud

- A 'Cloud' is an elastic execution environment of resources involving multiple stakeholders and providing a metered service at multiple granularities for a specified level of quality of service.

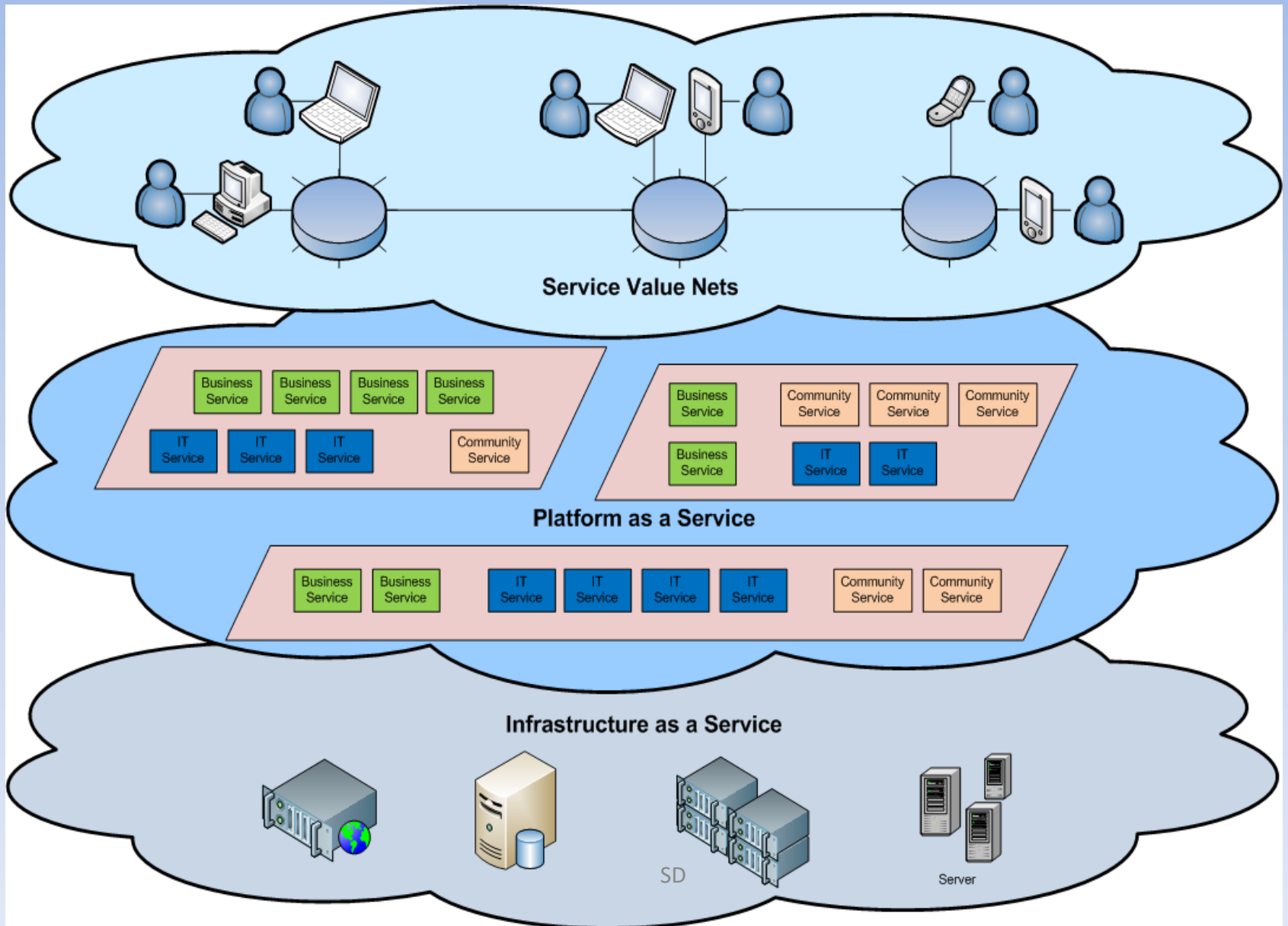
Typical File Fetching Procedure

- Process to store a file in a System
- Before fetching, we need to store
  - Open a document / file from USB or Memory
  - Copy the document / file into any drive in a PC
  - Open the document / file at any time from PC
  - Add administrative protocols for security
- What if, the above procedure is REMOTE?

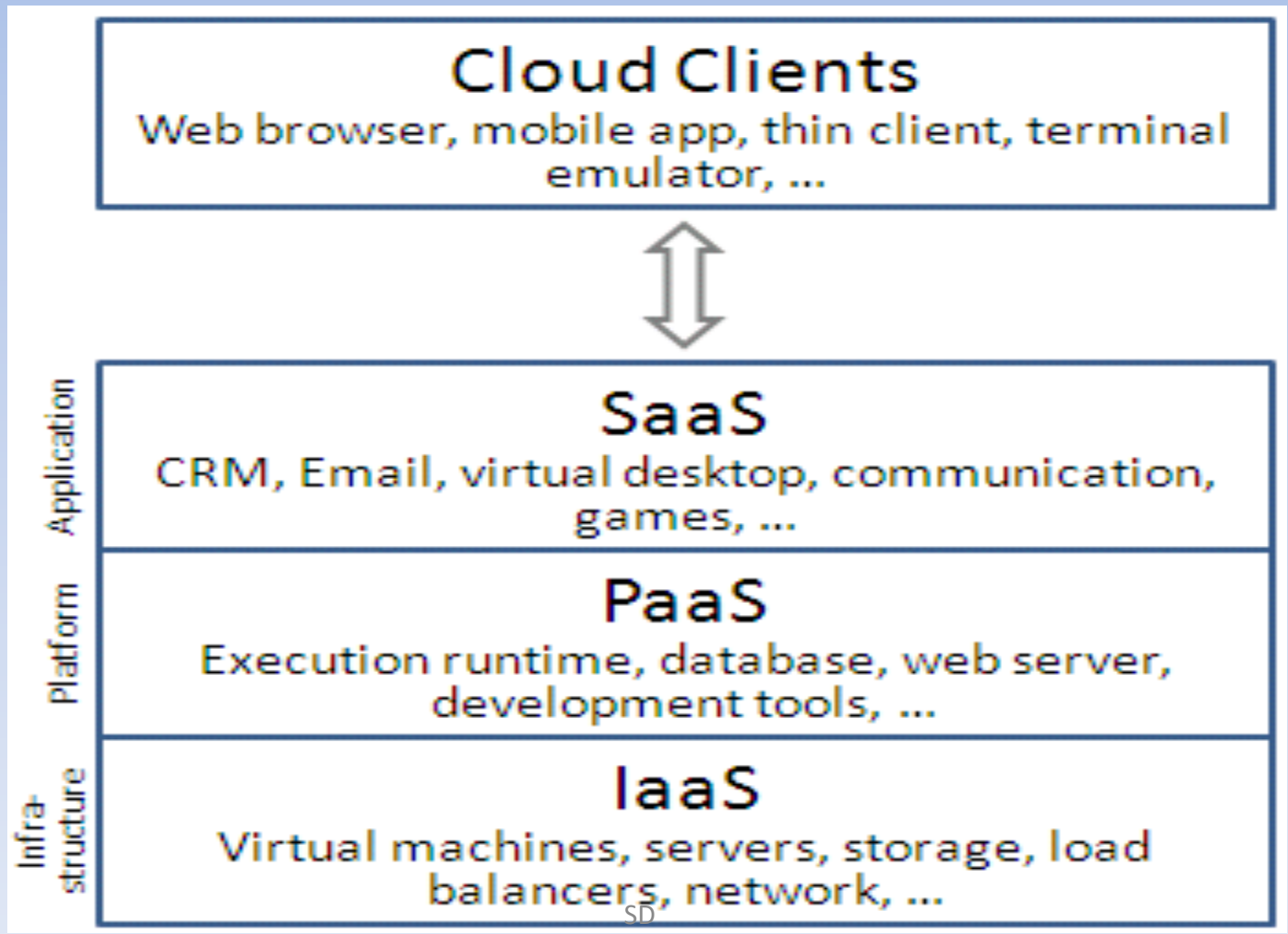
Types of Cloud

- Private Cloud
  - Specific to single organization
- Public Cloud
  - Anybody can use on pay-as-you-go model
- Community Cloud
  - Set of organizations based on policies
- Hybrid Cloud
  - Mixture of two or more clouds

Cloud Architecture



- Cloud Service Models
  - *IaaS, PaaS, SaaS* but now *XaaS*



Cloud Computing Characteristics

Common Characteristics:

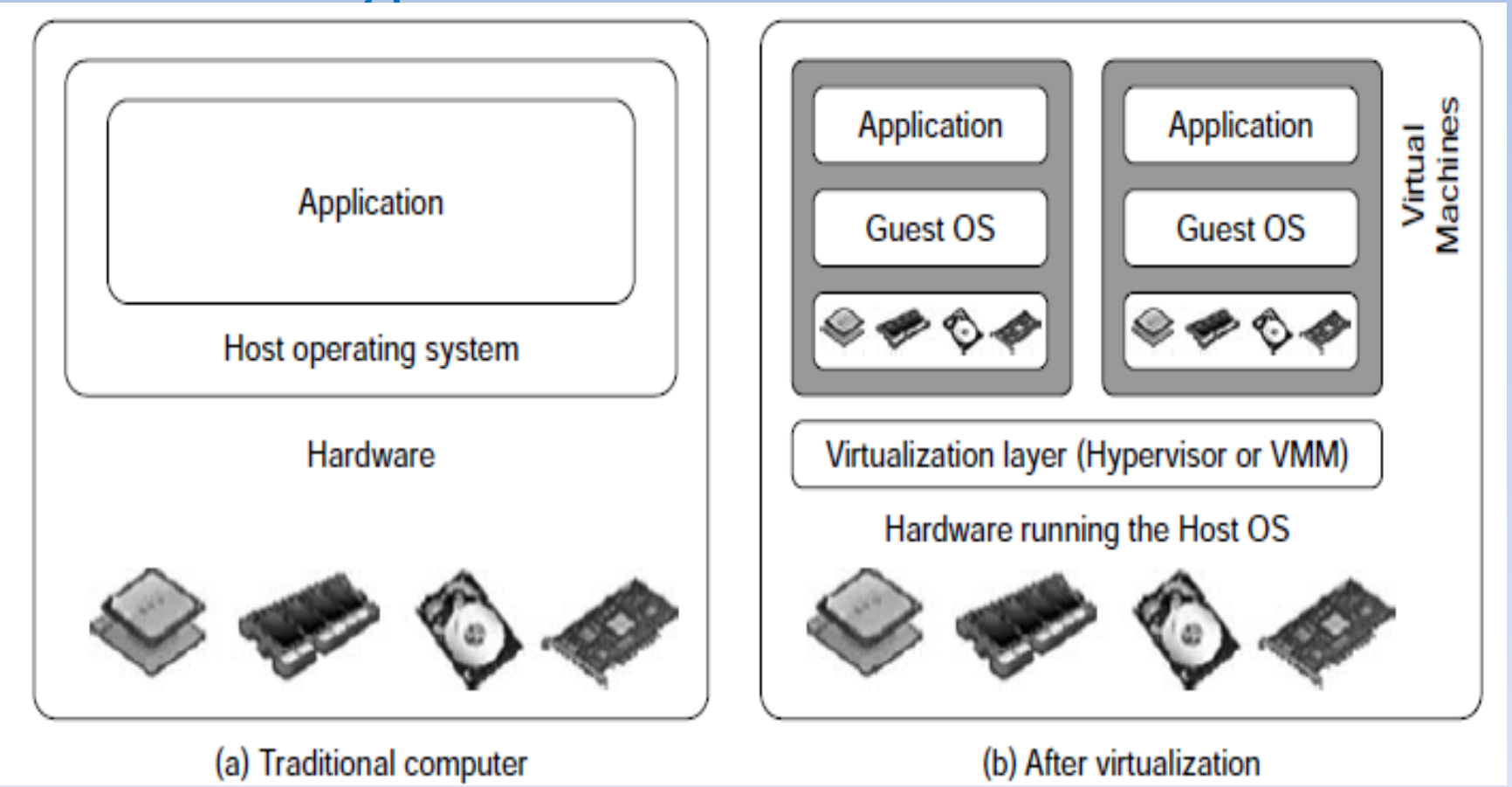
- Massive Scale
- Resilient Computing
- Homogeneity
- Geographic Distribution
- Virtualization
- Service Orientation
- Low Cost Software
- Advanced Security

Essential Characteristics:

- On Demand Self-Service
- Broad Network Access
- Rapid Elasticity
- Resource Pooling
- Measured Service

Role of VM and Hypervisor

- Importance of Virtual Machine (VM)
- Role of Hypervisor



Container

- A lightweight OS-level virtualization method along with the stand-alone piece of executable software and NOT a virtual machine. And it can also be called as process with isolation, shared resources, and layered filesystems

Terminologies in Cloud

- Data Center
- DNS
- Job Queue
- Auto Scaling
- Scheduling
- Load Balancing
- SLA : Service Level Agreements
- Execution of Cloudlet

AWS Reference Architectures

- Amazon EC2
- Amazon EC2 Spot
- Auto Scaling
- Elastic Load Balancing
- Amazon Route 53
- Amazon CloudFront
- Amazon S3
- Amazon RDS
- Amazon SQS

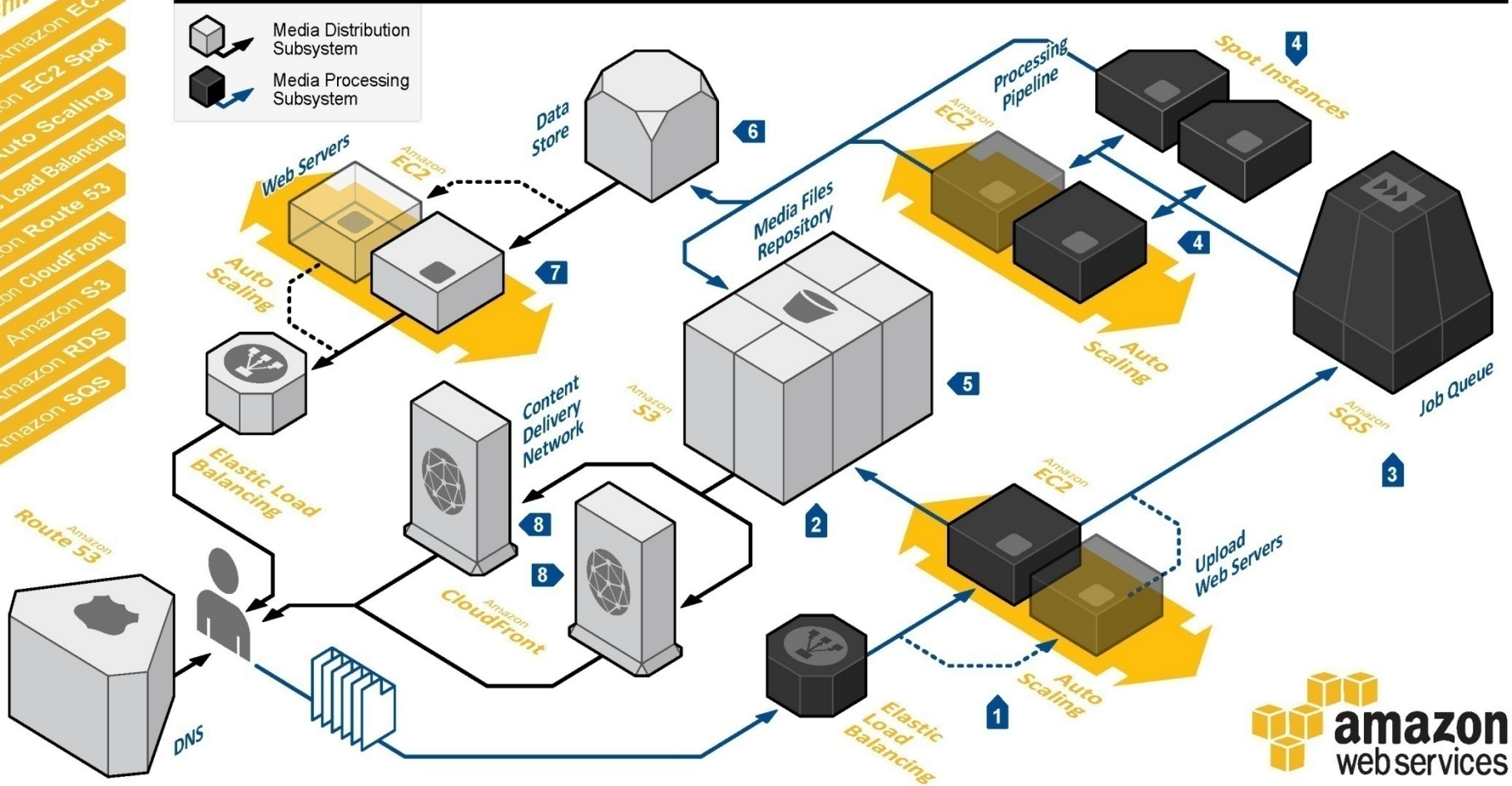
MEDIA SHARING

Media sharing is one of the hottest markets on the Internet. Customers have a staggering appetite for placing photos and videos on social networking sites, and for sharing their media in custom online photo albums.

The growing popularity of media sharing means scaling problems for site owners, who face ever-increasing storage and bandwidth requirements and increased go-to-market pressure to deliver faster than the competition.

Since most businesses today have limited manpower, budget, and data center space, AWS offers a unique set of opportunities to compete and scale without having to invest in hardware, staff, or additional data center space. Utilizing AWS is not an all or nothing proposition. Depending on the project, different services can be used independently.

This diagram shows an example of a highly available, durable, and cost-effective media sharing and processing platform.



System Overview

1 Sharing content first involves uploading media files to the online service. In this configuration, an **Elastic Load Balancer** distributes incoming network traffic to upload servers, a dynamic fleet of **Amazon Elastic Compute Cloud** (Amazon EC2) instances. **Amazon CloudWatch** monitors these servers and an **Auto Scaling** group manages them, automatically scaling EC2 capacity up or down based on load. In this example, a separate endpoint to receive media uploads was created in order to off-load this task from the website's servers.

2 Original uploaded files are stored in **Amazon Simple Storage Service** (Amazon S3), a highly available and durable storage service.

3 To submit a new file to be processed, upload web servers push a message into an **Amazon Simple Queue Service** (Amazon SQS) queue. This queue acts as a communication pipeline between the file reception and file processing components.

4 The processing pipeline is a dedicated group of **Amazon EC2** instances used to execute any kind of post-processing task on the uploaded media files (video transcoding, image resizing, etc.). To automatically adjust the needed capacity, **Auto Scaling** manages this group. You can use **Spot Instances** to dynamically extend the capacity of the group and to significantly reduce the cost of file processing.

5 Once processing is completed, **Amazon S3** stores the output files. Original files can be stored with high durability. Processed files could use reduced redundancy.

6 Media-related data can be put in a relational database like **Amazon Relational Database Service** (Amazon RDS) or in a key-value store like **Amazon SimpleDB**.

7 A third fleet of EC2 instances is dedicated to host the website front-end of the media sharing service.

8 Media files are distributed from **Amazon S3** to the end user via **Amazon CloudFront**, a content delivery network. **Amazon CloudFront** offers low-latency delivery through a worldwide network of edge locations.

Advantages and QoS Parameters

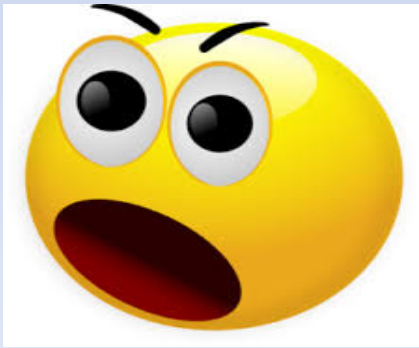
- Advantages of Cloud
  - *No up-front investment*
  - *Lowering operating cost*
  - *Highly scalable*
  - *Easy access*
  - *Reducing business risks and maintenance expenses*
- QoS parameters
  - *Resource Utilization, Waiting Time, Throughput, Migration Time, Execution Time, Reliability, Performance, Security.*

Cloud Simulators

- CloudSim : From University of Melbourne, Australia. Its open source
  - Teach Cloud, Ground Sim, DC Sim
- iCanCloud : Has graphical user interface
- SPECI : Simulation Program for Elastic Cloud Infrastructure
- GreenCloud
- Open Stack

Thoughts on Cloud w.r.t Industry

- More application availability on the cloud
  - *With most new software being built for cloud from the outset, it is predicted that by 2016 over a quarter of all applications (around 48 million) will be available on the cloud.*
- Increased growth in the market for cloud
  - *According to Gartner cloud is growing from $18.2 billion in 2012 to $>100 billion in 2020.*
- More innovation because of cloud



Issues In Cloud Computing

- Security and Privacy
- Performance
- Reliability and Availability
- Scalability and Elasticity
- Interoperability and Portability
- Resource Management and Scheduling
- Load Balancing
- Energy Consumption
- VM Migration