The Business of Cloud

Dr. Deepak Saxena, SME IIT Jodhpur

Cloud Provider Perspective

- Building cloud environment
  - OS, Memory, Processing, Backup capabilities

- Monitoring
  - Virtual server lifecycle, Data storage, Network traffic, Failure conditions, Event triggers

- Security
  - Encryption, hashing, digital signature, Identity and Access Management (IAM), SSO (Single sign on)

- Service
  - Service load balancing, Dynamic failure detection and recovery, Storage maintenance, Elastic resource capability

Cloud Consumer Perspective (IaaS)

- Controlling scalability features (automated scaling, load balancing)
- Controlling the lifecycle of virtual IT resources (shutting down, restarting, powering up of virtual devices)
- Controlling the virtual network environment and network access rules (firewalls, logical network perimeters)
- Establishing and displaying service provisioning agreements (account conditions, usage terms)
- Managing the pre-allocation of cloud-based IT resources (resource reservation)
- Managing credentials and passwords for cloud resource administrators
- Managing credentials for cloud-based security groups that access virtualized IT resources through an IAM
- Managing security-related configurations
- Managing customized virtual server image storage (importing, exporting, backup)
- Selecting high-availability options (failover, IT resource clustering)
- Selecting and monitoring SLA metrics
- Selecting basic software configurations (operating system, pre-installed software for new virtual servers)
- Selecting IaaS resource instances from a number of available hardware-related configurations and options (processing capabilities, RAM, storage)
- Selecting the geographical regions in which cloud-based IT resources should be hosted.
- Tracking and managing costs

Cloud Consumer Perspective (PaaS)

- Establishing and displaying service provisioning agreements, such as account conditions and usage terms
- Selecting software platform and development frameworks for ready-made environments
- Selecting instance types, which are most commonly frontend or backend instances
- Selecting cloud storage devices for use in ready-made environments
- Controlling the lifecycle of PaaS-developed applications (deployment, starting, shutdown, restarting, and release)
- Controlling the versioning of deployed applications and modules
- Configuring availability and reliability-related mechanisms
- Managing credentials for developers and cloud resource administrators using IAM
- Managing general security settings, such as accessible network ports
- Selecting and monitoring PaaS-related SLA metrics
- Managing and monitoring usage and IT resource costs
- Controlling scalability features such as usage quotas, active instance thresholds, and the configuration and deployment of the automated scaling
- Listener and load balancer mechanisms

Cloud Consumer Perspective (SaaS)

- Managing security-related configurations

- Managing select availability and reliability options

- Managing usage costs

- Managing user accounts, profiles, and access authorization

- Selecting and monitoring SLAs

- Setting manual and automated scalability options and limitations
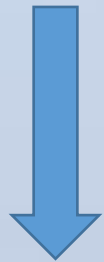
Business Cost Metrics

On-Premise

- Up-Front
  - Hardware
  - Software
  - Labor required for deployment.

- Ongoing
  - Licensing fees
  - Electricity
  - Insurance
  - Labor

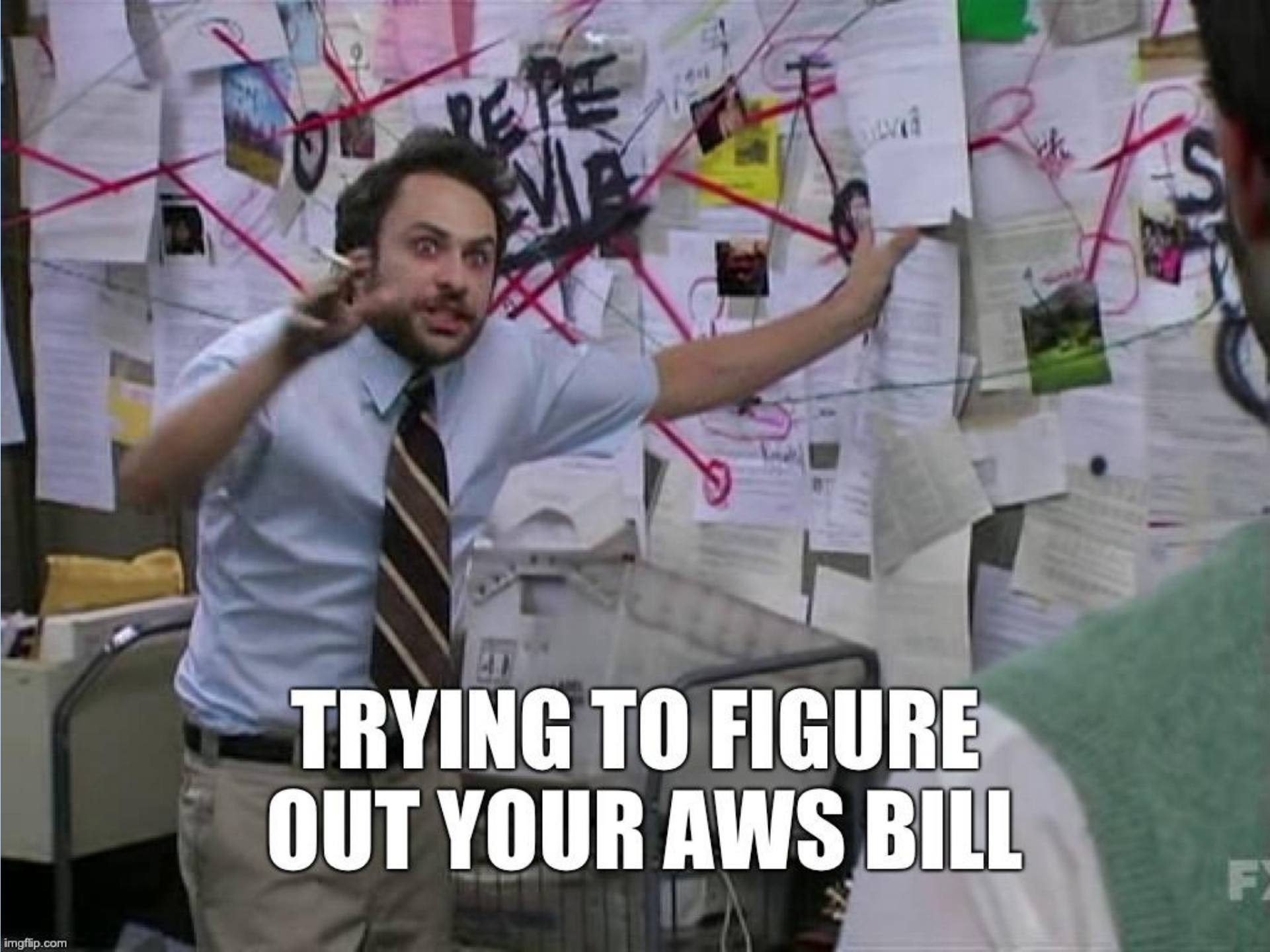**Cloud**

- Up-Front
  - labor costs required to assess and set up a cloud environment.

- Ongoing
  - Virtual hardware leasing fees,
  - Bandwidth usage fees
  - Licensing fees
  - Labor.

TRYING TO FIGURE OUT YOUR AWS BILL

Some sample calculations

Up-Front Costs	Cloud Environment	On-Premise Environment
Hardware	$0	$15,000
Licensing	$0	$30,500
Labor	$5,000	$5,500
Total Up-Front Costs	**$5,000**	**$51,000**

Monthly On-Going Costs	Cloud Environment	On-Premise Environment
Application Servers	$2,070	$0
Database Servers	$327	$0
WAN Network	$42	$0
Environment	$0	$750
Software Licensing	$0	$520
Hardware Maintenance	$0	$100
Administration	$800	$2,600
Total On-Going Costs	**$3,139**	**$3,970**

In this case, company may decide moving to cloud.

Some sample calculations

Up-Front Costs	Cloud Environment	On-Premise Environment
Hardware	$0	$15,000
Licensing	$0	$15,200
Labor	$45,000	$5,500
Total Up-Front Costs	**$45,000**	**$35,700**

Monthly On-Going Costs	Cloud Environment	On-Premise Environment
Application Servers	$3,420	$0
Database Servers	$1,635	$0
WAN Network	$80	$0
Environment	$0	$1,050
Software Licensing	$0	$300
Hardware Maintenance	$0	$100
Administration	$1,200	$4,500
Total On-Going Costs	**$6,335**	**$5,950**

In this case, company may decide not moving to cloud.

Cloud Usage Cost Metrics

- Network usage
- Server usage
- Cloud storage device
- Cloud service

Network Usage

- Inbound
  - *Description* – inbound network traffic
  - *Measurement* – Σ, inbound network traffic in bytes
  - *Frequency* – continuous and cumulative over a predefined period
  - *Cloud Delivery Model* – IaaS, PaaS, SaaS
  - *Example* – up to 1 GB free, $0.001/GB up to 10 TB a month

- Outbound
  - *Description* – outbound network traffic
  - *Measurement* – Σ, outbound network traffic in bytes
  - *Frequency* – continuous and cumulative over a predefined period
  - *Cloud Delivery Model* – IaaS, PaaS, SaaS
  - *Example* – up to 1 GB free a month, $0.01/GB between 1 GB to 10 TB per month

- Intra-cloud

Network Usage

- Intra-cloud
    - *Description* – network traffic between geographically diverse IT resources of the same cloud
    - *Measurement* – Σ, intra-cloud WAN traffic in bytes
    - *Frequency* – continuous and cumulative over a predefined period
    - *Cloud Delivery Model* – IaaS, PaaS, SaaS

    - *Example* – up to 500 MB free daily and $0.01/GB thereafter, $0.005/GB after 1 TB per month

Server Usage

- ## On-demand Virtual Machine Instance Allocation

  - *Description* – uptime of a virtual server instance
  - *Measurement* – Σ, virtual server start date to stop date
  - *Frequency* – continuous and cumulative over a predefined period
  - *Cloud Delivery Model* – IaaS, PaaS

  - *Example* – $0.10/hour small instance, $0.20/hour medium instance, $0.90/hour large instance

- ## Reserved Virtual Machine Instance Allocation

  - *Description* – up-front cost for reserving a virtual server instance

  - *Measurement* – Σ, virtual server reservation start date to expiry date

  - *Frequency* – daily, monthly, yearly

  - *Cloud Delivery Model* – IaaS, PaaS

  - *Example* – $55.10/small instance, $99.90/medium instance, $249.90/large instance

Cloud Storage device Usage

- ## On-demand storage space allocation

  - *Description* – duration and size of on-demand storage space allocation in bytes
  - *Measurement* – Σ, date of storage release / reallocation to date of storage allocation (resets upon change in storage size)
  - *Frequency* – continuous
  - *Cloud Delivery Model* – IaaS, PaaS, SaaS

- ## I/O Data Transferred Metric

  - *Description* – amount of transferred I/O data
  - *Measurement* – Σ, I/O data in bytes
  - *Frequency* – continuous
  - *Cloud Delivery Model* – IaaS, PaaS
  - *Example* – $0.10/TB

Cloud Service Usage

- **Application Subscription Duration Metric**
  - *Description* – duration of cloud service usage subscription
  - *Measurement* – Σ, subscription start date to expiry date
  - *Frequency* – daily, monthly, yearly
  - *Cloud Delivery Model* – SaaS
  - *Example* – $69.90 per month

- **Number of Nominated Users Metric**
  - *Description* – number of registered users with legitimate access
  - *Measurement* – number of users
  - *Frequency* – monthly, yearly
  - *Cloud Delivery Model* – SaaS
  - *Example* – $0.90/additional user per month

- **Number of Transactions Users Metric**
  - *Description* – number of transactions served by the cloud service
  - *Measurement* – number of transactions (request-response message exchanges)
  - *Frequency* – continuous
  - *Cloud Delivery Model* – PaaS, SaaS
  - *Example* – $0.05 per 1,000 transactions

Pricing as per model

- IaaS

  Pricing is usually based on IT resource allocation and usage, which includes the amount of transferred network data, number of virtual servers, and allocated storage capacity.

- PaaS

  Similar to IaaS, this model typically defines pricing for network data transferred, virtual servers, and storage. Prices are variable depending on factors such as software configurations, development tools, and licensing fees.

- SaaS

  Because this model is solely concerned with application software usage, pricing is determined by the number of application modules in the subscription, the number of nominated cloud service consumers, and the number of transactions.

Exact pricing model depends upon many considerations

- Market competition and regulatory requirements

- Overhead incurred during the design, development, deployment, and operation of cloud services and other IT resources

- Opportunities to reduce expenses via IT resource sharing and data center optimization

- Negotiation

Service Level Agreement (SLA)

- SLAs issued by cloud providers are human-readable documents that describe quality-of-service (QoS) features, guarantees, and limitations of one or more cloud-based IT resources.

- SLAs use service quality metrics (also called SLI – Service Level Indicators) to express measurable QoS characteristics.
  - *Availability* – up-time, outages, service duration
  - *Reliability* – minimum time between failures, guaranteed rate of successful responses
  - *Performance* – capacity, response time, and delivery time guarantees
  - *Scalability* – capacity fluctuation and responsiveness guarantees
  - *Resiliency* – mean-time to switchover and recovery

Service Availability Metric

- Availability Rate Metric
  - *Description* – percentage of service up-time
  - *Measurement* – total up-time / total time
  - *Frequency* – weekly, monthly, yearly
  - *Cloud Delivery Model* – IaaS, PaaS, SaaS
  - *Example* – minimum 99.5% up-time

- Outage Duration Metric
  - *Description* – duration of a single outage
  - *Measurement* – date/time of outage end – date/time of outage start
  - *Frequency* – per event
  - *Cloud Delivery Model* – IaaS, PaaS, SaaS
  - *Example* – 1 hour maximum, 15 minute average

Service Reliability Metric

- ## Mean-Time Between Failures (MTBF) Metric

  - *Description* – expected time between consecutive service failures
  - *Measurement* – Σ (start of downtime – start of uptime) / number of failures
  - *Frequency* – monthly, yearly
  - *Cloud Delivery Model* – IaaS, PaaS
  - *Example* – 90 day average

- ## Reliability Rate Metric

  - *Description* – percentage of successful service outcomes under pre-defined conditions
  - *Measurement* – total number of successful responses / total number of requests
  - *Frequency* – weekly, monthly, yearly
  - *Cloud Delivery Model* – SaaS
  - *Example* – minimum 99.5%

Service Performance Metrics

Metric	Definition	Applicable Models	Frequency	Example
Network Capacity	measurable characteristics of network capacity	IaaS, PaaS, SaaS	Continuous	10 mbps
Storage Device Capacity	measurable characteristics of storage device capacity	IaaS, PaaS, SaaS	Continuous	80 GB of storage
Server Capacity	measurable characteristics of server capacity	IaaS, PaaS	Continuous	1 core at 1.7 GHz, 16 GB of RAM, 80 GB of storage
Web Application Capacity	measurable characteristics of Web application capacity	SaaS	Continuous	Maximum 100,000 requests per minute
Instance Starting Time	length of time required to initialize a new instance	IaaS, PaaS	Per event	5 minute maximum, 3 minute average
Response Time	time required to perform synchronous operation	SaaS	daily, weekly, monthly	5 millisecond average
Completion Time	time required to complete an asynchronous task	PaaS, SaaS	daily, weekly, monthly	1 second average

Service Scalability Metrics

Metric	Definition	Applicable Models	Frequency	Example
Storage Scalability (Horizontal)	permissible storage device capacity changes in response to increased workloads	IaaS, PaaS, SaaS	Continuous	1,000 GB maximum (automated scaling)
Server Scalability (Horizontal)	permissible server capacity changes in response to increased workloads	IaaS, PaaS	Continuous	1 virtual server minimum, 10 virtual server maximum (automated scaling)
Server Scalability (Vertical)	permissible server capacity fluctuations in response to workload fluctuations	IaaS, PaaS	Continuous	512 core maximum, 512 GB of RAM

Service Resiliency Metrics

Metric	Definition	Applicable Models	Frequency	Example
Mean-Time to Switchover (MTSO)	the time expected to complete a switchover from a severe failure to a replicated instance in a different geographical area	IaaS, PaaS, SaaS	monthly, yearly	10 minutes average
Mean-Time System Recovery (MTSR)	time expected for a resilient system to perform a complete recovery from a severe failure	IaaS, PaaS, SaaS	monthly, yearly	120 minutes average

Some Example SLAs

- AWS
- Google Cloud Platform
- Azure
- IBM