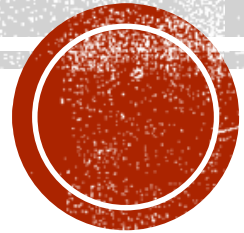
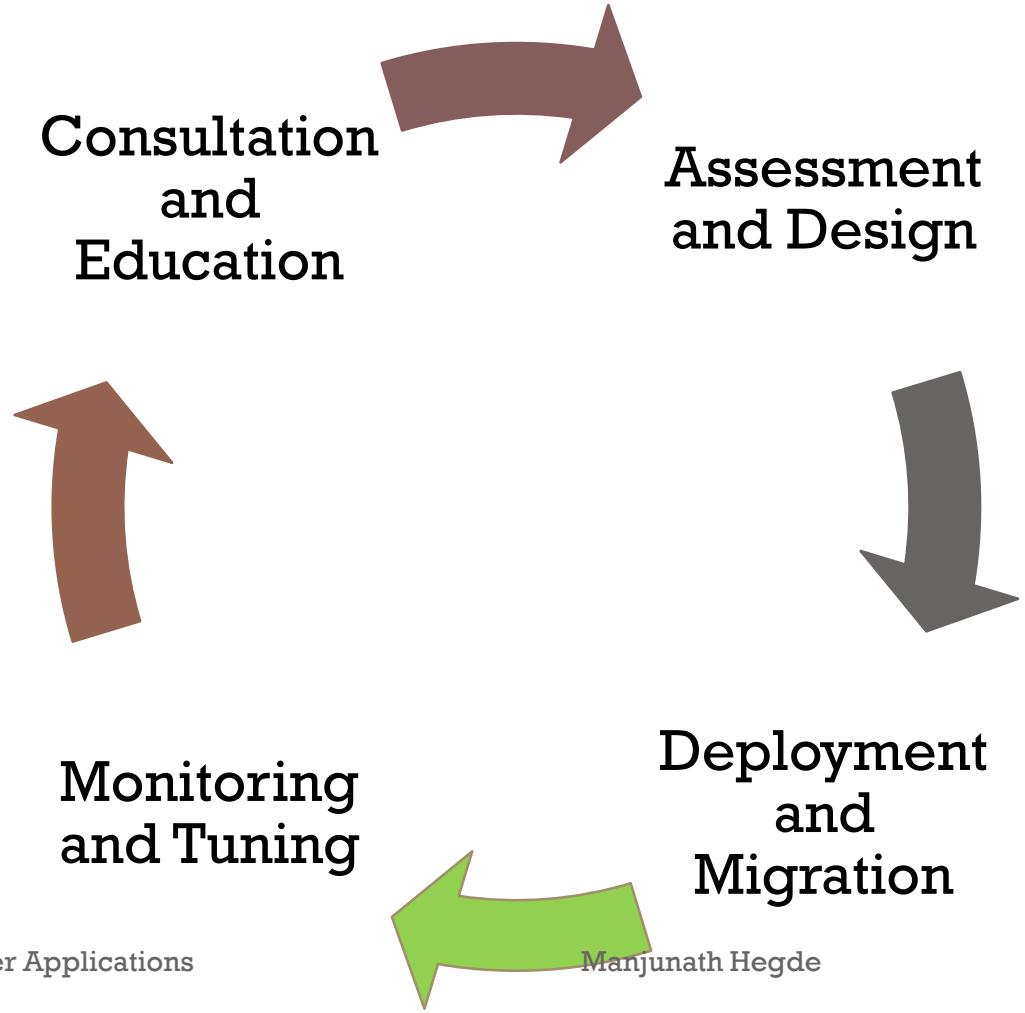
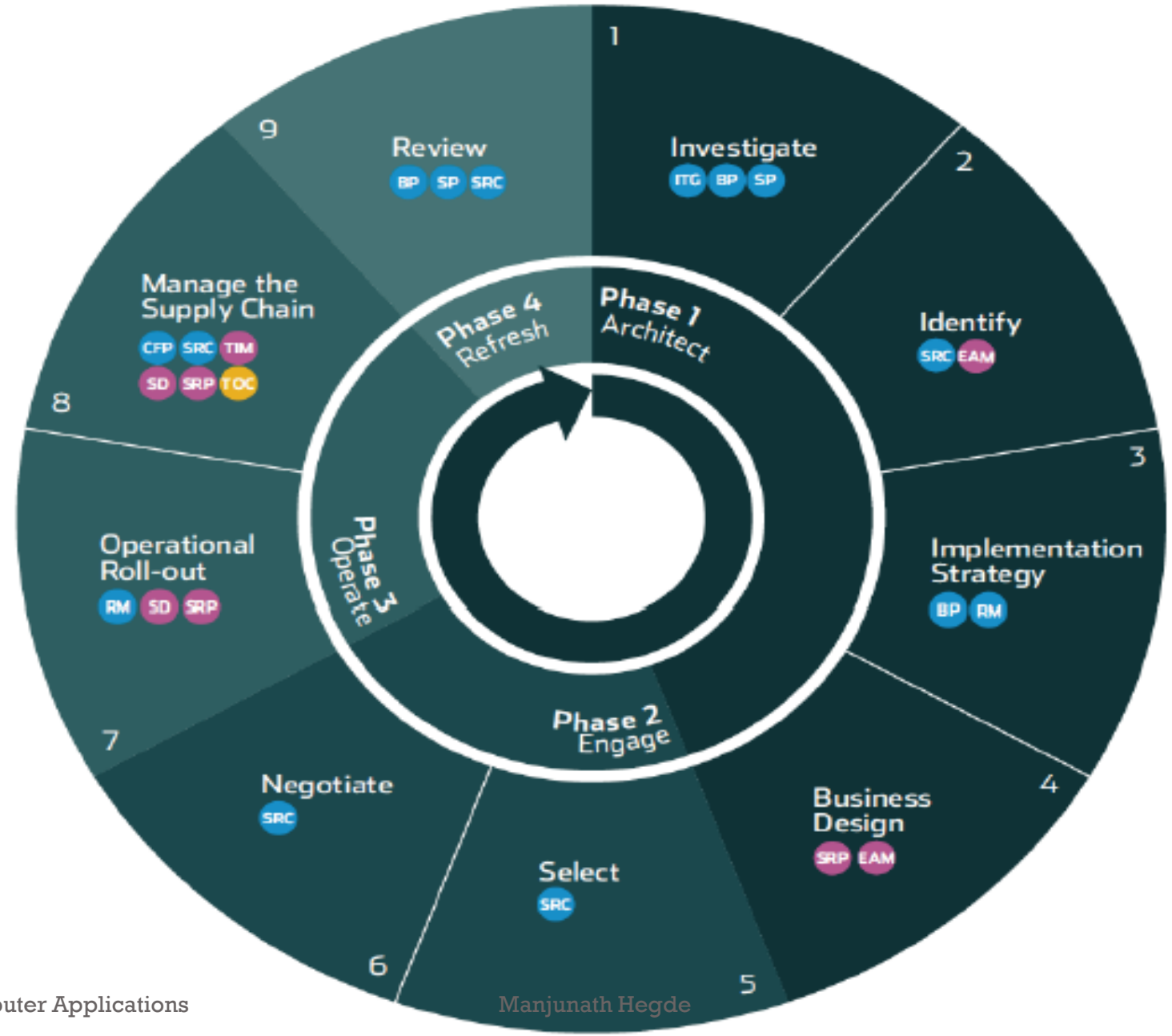
CLOUD SELECTION & LIFECYCLE



Dr. Manjunath Hegde
Dept. of Computer Applications
Manipal Institute of Technology, Manipal

CLOUD COMPUTING LIFE CYCLE




1
Investigate
ITG BP SP
2
Identify
SRC EAM
3
Implementation Strategy
BP RM
4
Business Design
SRP EAM
5
Select
SRC
6
Negotiate
SRC
7
Operational Roll-out
RM SD SRP
8
Manage the Supply Chain
CFP SRC TIM
SD SRP TOC
9
Review
BP SP SRC
Phase 1 Architect
Phase 2 Engage
Phase 3 Operate
Phase 4 Refresh

CONT..

- Architect: Starts with the investigation and planning of the cloud project. An organization will only commit a small number of high-level resources in order to decide if they should go ahead with a full-scale project.
- Engage: Selects a service provider that can deliver the required cloud service.
- Operate: Implementation and the day-to-day management of the cloud Service.
- Refresh: Ongoing review of cloud services.

INVESTIGATE

Activities	Outputs
• Determine the organization's IT objectives and its alignment with the business. • Determine what role cloud computing will play within the IT strategy. • Gather intelligence on cloud service offerings. • Validate with cloud subject matter experts.	• IT strategy for cloud computing. • Strategic intent of moving to the cloud and how it progresses the business objectives. • Intelligence document on cloud service offerings and providers. • Documented understanding of what will be achieved by comparing the strategic requirements with the available services and providers.

IDENTIFY

Activities	Outputs
• Determine what services will be outsourced to the cloud, and consider impacts on the service, people, cost, infrastructure, and stakeholders. • Decide what type of cloud outsourcing model will be used, and why it is suitable. • Document the current and future states of the IT infrastructure.	• A List of services to be outsourced to the cloud, with documented understanding on impacts to service, people, cost, infrastructure, and stakeholders. • A Cloud outsourcing model, with documented justification. • Documented current and future states of the IT structure.

IMPLEMENTATION STRATEGY

Activities	Outputs
• Determine the roll-out approach and how the program will be managed. • Detail how the program will be staffed and reported. • Decide how cloud suppliers will be engaged, selected and managed. • Determine how risks will be assessed and managed, including data recovery and in-sourcing.	• A program roll-out strategy. • A Communication strategy. • A strategy to manage staff impacted by the migration to cloud. • A Cloud risk management strategy. • A Cloud supplier management strategy.

BUSINESS DESIGN

Activities	Outputs
• Detail the service offering you wish to tender for. • Clearly define negotiable / non-negotiable issues around contracts, service-level agreements (SLA), and pricing model	• Detailed and clear tender documents for cloud suppliers.

SELECTION

Activities	Outputs
• Define the tender/bid process. • Select and staff an evaluation team. • Invite bids/tenders. • Evaluate suppliers against the defined criteria. • Shortlist the supplier(s). • Carry out due diligence.	• A tender process. • Evaluation criteria. • A shortlist of suitable suppliers with caveats. • A Due diligence report.

NEGOTIATE

Activities	Outputs
• Define the negotiation strategy. • Select and staff the negotiation team. • Carry out negotiations. • Select the preferred cloud supplier. • Get internal approvals and sign the contract.	• A negotiation strategy. • Results of the negotiation. • Signed final documents: Contract, SLA and Pricing document.

OPERATIONAL ROLL OUT

Activities	Outputs
• Finalize and publish transition plans. • Select and staff the transition team. • Agree and publish acceptance criteria. • Carry out the transition. • Communicate progress. • Conduct knowledge transfer. • Manage staff (directly and indirectly) impacted.	• A roll-out plan. • Progress updates. • A signed acceptance document.

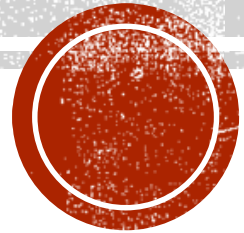
MANAGE SUPPLY CHAIN

Activities	Outputs
• Manage and report at cloud service operational level. • Capture and manage issues, variations and disputes. • Manage the supplier relationship. • Change management. • Continuous improvement. • Assess and validate how the cloud service is performing.	• Day-to-day cloud service performance metrics. • Status on issues, problems, variations, and disputes. • Supplier meeting minutes. • A change management report. • Audit reports.

REVIEW

Activities	Outputs
• Gather intelligence on the relevant market segment, cloud service technology trends, and supplier offerings. • Audit cloud supplier performance and compare to alternatives. • Understand and assess how other changes in the organization impact on the existing cloud service arrangement. • Based on the above inputs, regularly reassess and review requirements. • Make and present a business case for any significant change to the current cloud service arrangement in order to get approval to start a new cycle.	• An intelligence report for next generation cloud service offerings. • Cloud supplier audit results. • A business case for any proposed changes.

SERVICE LEVEL AGREEMENTS (SLAS)



INTRODUCTION

- A service-level agreement (SLA) is defined as an official commitment that prevails between a service provider and the customer.
- Particular aspects of the service – quality, availability, responsibilities – are agreed between the service provider and the service user.
- The most common component of SLA is that the services should be provided to the customer as agreed upon in the contract.
- As an example, Internet service providers and telcos will commonly include service level agreements within the terms of their contracts with customers to define the level(s) of service being sold in plain language terms.
- In this case the SLA will typically have a technical definition in terms of mean time between failures (MTBF), mean time to repair or mean time to recovery (MTTR).

SERVICE LEVEL AGREEMENT CONSIDERATIONS

- Use of data/Security
- Location of data
- No change of terms
- Destruction
- Ownership (assignment)
- Subpoena response
- Regulatory requirements
- Insurance/Indemnity
- Audits

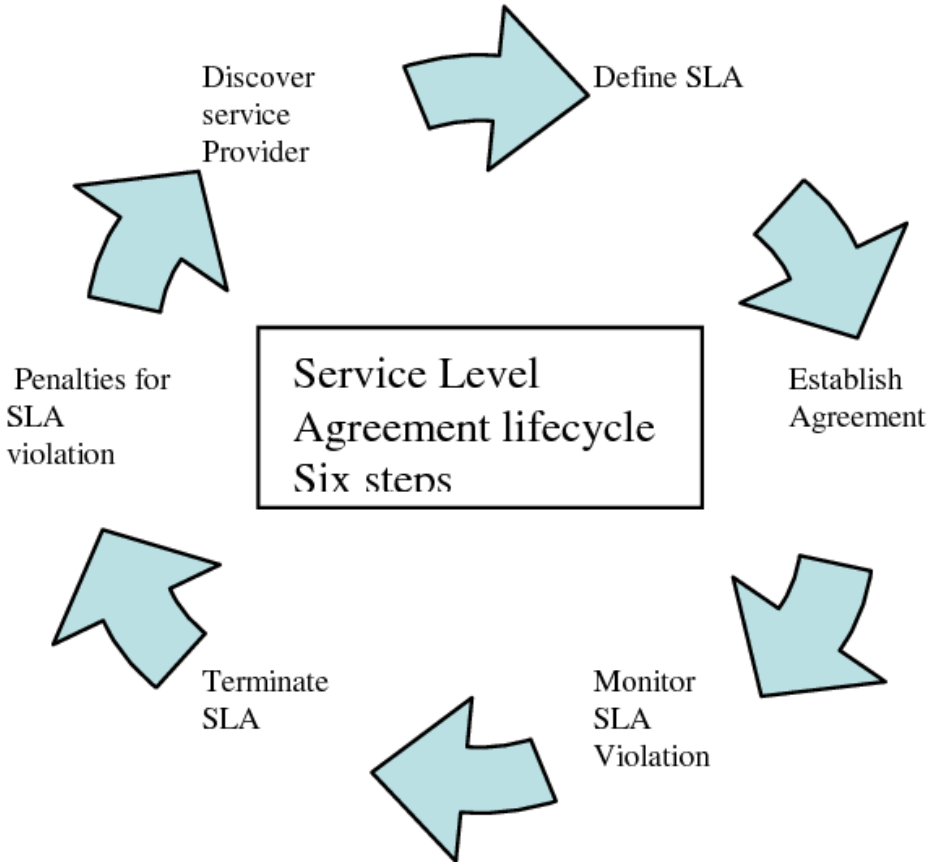
PC: https://www.trinityp3.com/wp-content/uploads/2008/08/SLA_Service_Level_Agreement1.jpg

SLA CONTAINS

SLA should contain:

- The list of services the provider will deliver and a complete definition of each service.
- Metrics to determine whether the provider is delivering the service as promised
- Auditing mechanism to monitor the service.
- Responsibilities of the provider and the consumer
- Remedies available to both provider and client if the terms of the
- SLA are not met.
- A description of how the SLA will change over time.

SAL LIFECYCLE



PHASES OF CREATING SLA

- SLA Development in which the SLA templates are developed,
- Negotiation and Sales: In this phase the SLA is negotiated and the contracts are executed,
- Implementation: where the SLA is generated.
- Execution: The SLA is executed, monitored, and maintained,
- Assessment: Evaluation of the SLA performance. In this phase, a re-evaluation of the initial SLA template might be done

SLAS CONCERNS (1/3)

- Security: Client and Cloud Service Provider (CSP) must understand security requirements.
- Data encryption: Data must be encrypted while it is in motion and while it is at rest. The details of the encryption algorithms and access control policies should be specified.
- Privacy: Basic privacy concerns are addressed by requirements such as data encryption, retention, and deletion. An SLA should make it clear how the cloud provider isolates data and applications in a multi-tenant environment.
- Data retention/deletion: How does CSP prove they comply with retention laws and deletion policies?
- Hardware erasure/ destruction: Same as #4.

SLAS CONCERNS (2/3)

- Regulatory compliance: If regulations must be enforced because of the type of data, CSP must be able to prove compliance.
- Transparency: For critical data and applications CSP must be proactive in notifying client when the terms of the SLA are breached including infrastructure issues like outages and performance problems as well as security incidents.
- Certification: CSP should be responsible for proving required certification and keeping it current.
- Performance definitions: Defining terminology such as uptime and other contractual metric terms (i.e. – uptime could mean all servers on continent are available or only one designated server is available.)

SLAS CONCERNS (3/3)

- Monitoring: Responsible party for monitoring including identification of any third-party organization designated to monitor performance of the provider.
- Audit Rights: To monitor for any data breaches including loss of data and availability issues. SLA should clarify when and how the audits will take place.
- Metrics: to be monitored in real-time and audited after occurrence. Metrics of an SLA must be objectively and unambiguously defined.
- Human interaction: On-demand self-service is one of the basic characteristics of cloud computing, but SLA should provide customer service when needed.

CONTRACT ISSUE - IN REALITY

- Currently, the standard contracts offered by cloud computing providers are one-sided and service provider-friendly, with little opportunity to change terms.
- Few offer meaningful service levels or assume any responsibility for legal compliance, security or data protection. Many permit suspension of service or unilateral termination, and disclaim all or most of the provider's potential liability.
- In addition, some cloud computing providers emphasize low cost offerings, which leave little room for robust contractual commitments or customer requirements.

SERVICE-LEVEL OBJECTIVE (SLO)

- A service-level objective (SLO) is a key element of a service-level agreement (SLA) between a service provider and a customer.
- SLOs are agreed upon as a means of measuring the performance of the Service Provider and are outlined as a way of avoiding disputes between the two parties based on misunderstanding.
- There is often confusion in the use of SLAs and SLOs.
- The SLA is the entire agreement that specifies what service is to be provided, how it is supported, times, locations, costs, performance, and responsibilities of the parties involved.
- SLOs are specific measurable characteristics of the SLA such as availability, throughput, frequency, response time, or quality.
- These SLOs together are meant to define the expected service between the provider and the customer and vary depending on the service's urgency, resources, and budget.
- SLOs provide a quantitative means to define the level of service a customer can expect from a provider.

SCALABILITY AND LOCK-IN

SCALABILITY

- **Capability of a system, network, or process to handle growing amount of work, or its potential to be enlarged in order to accommodate that growth.**
- A system whose performance improves after adding hardware/capacity, is said to be a scalable system.
- An algorithm, design, networking protocol, program, or other system is said to scale if it is suitably efficient and practical when applied to large situations.
- If the design or system fails when a quantity increases, it does not scale.

DIMENSIONS OF SCALABILITY

- **Administrative scalability:** Ability for an increasing number of organizations or users to easily share single distributed system.
- **Functional scalability:** Ability to enhance the system by adding new functionality at minimal effort.
- **Geographic scalability:** Ability to maintain performance, usefulness, or usability regardless of expansion from concentration in a local area to a more distributed geographic pattern.
- **Load scalability:** Ability for distributed system to easily expand & contract its resource pool to accommodate heavier or lighter loads or number of inputs. i.e, the ease with which system or component can be modified, added, or removed, to accommodate changing load.
- **Generation scalability:** Ability of system to scale up using new generations of components.
- **Heterogeneous Scalability:** Ability to use components from different vendors.

PERFORMANCE RELATED SCALABILITY

With regards to high performance computing there are two common notions of scalability:

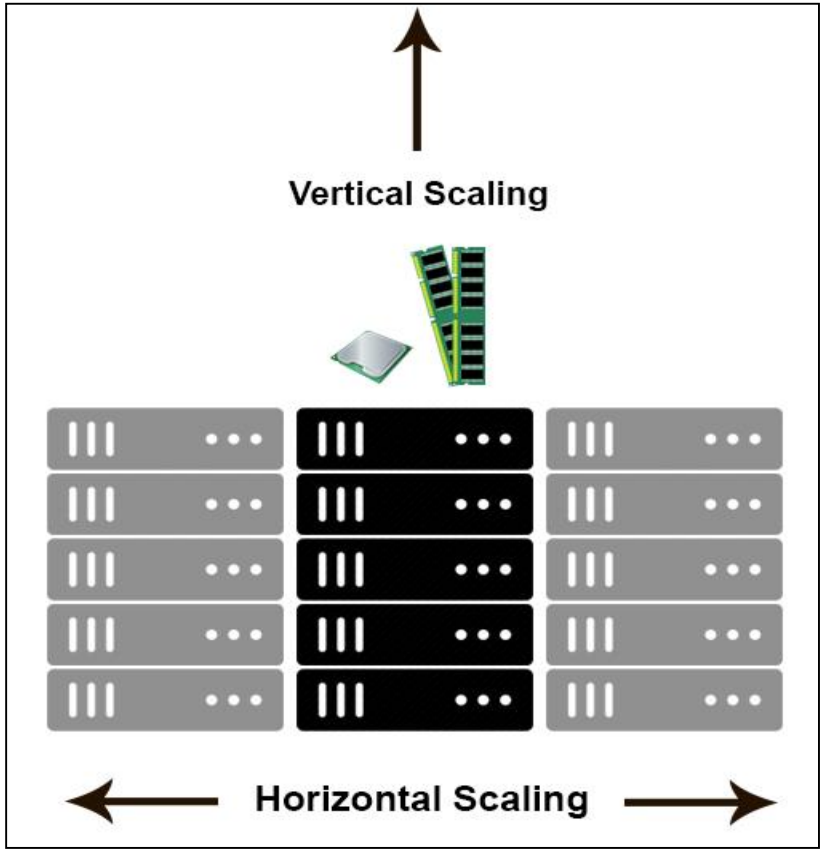
- **Strong scaling:** how the solution time varies with the number of processors for a fixed total problem size.
- **Weak scaling:** how the solution time varies with the number of processors for a fixed problem size per processor.

SCALABILITY TYPES

- Scaling is growing an infrastructure (compute, storage, networking) larger so that the applications riding on that infrastructure can serve more people at a time.
- The way in which the infrastructure is grown,
  - Scale Up (Vertical Scaling)
  - Scale Out (Horizontal Scaling)
  - Scaling Down
  - Scale In

VERTICAL & HORIZONTAL SCALING

- **Horizontal scaling:** Scale by adding more machines into the pool of resources.
- **Vertical scaling:** Scale by adding more power (CPU, RAM) to an existing machine.
- A machine on a server rack; and add more machines across horizontal direction, and add more resources to a machine in vertical direction



SCALE UP (VERTICAL SCALING)

- To add resources to single node in a system.
  - Addition of CPUs or memory to a single computer.
- Replacing what is existing with something more powerful.
- Pros
  - Less power consumption than running multiple servers
  - Cooling costs are less than scaling horizontally
  - Generally less challenging to implement (quantum-wise)
  - Less licensing costs
  - (sometimes) uses less network hardware than scaling horizontally
  - Easy to backup
- Cons
  - PRICE
  - Greater risk of hardware failure causing bigger outages
  - Generally severe vendor lock-in and limited upgradeability in the future

SCALE OUT (HORIZONTAL SCALING)

- Add more nodes to a system.
  - Adding more servers with less processors and RAM.
- Usually cheaper overall and can literally scale infinitely.
- Pros
  - Much cheaper than scaling vertically
  - Easier to run fault-tolerance
  - Easy to upgrade
- Cons
  - More licensing fees
  - Bigger footprint in the Data Center
  - Higher utility cost (Electricity and cooling)
  - Possible need for more networking equipment (switches/routers)

SCALING

- In database related environment; horizontal-scaling is often based on partitioning of data.
  - Each node contains only part of data .
- In vertical-scaling, data resides on a single node and scaling is done through multi-core.
  - Spreading the load between the CPU and RAM resources of that machine.
- Horizontal-scaling it is often easier to scale dynamically by adding more machines into the existing pool.
- Vertical-scaling is often limited to the capacity of a single machine.
- Scaling beyond that capacity often involves downtime, and comes with an upper limit.

SCALING

- Vertical scaling can essentially resize your server with no change to your code.
- Ability to increase the capacity of existing hardware or software by adding resources.
- Vertical scaling is limited by the fact that one can only get as big as the size of the server.

- Horizontal scaling affords the ability to scale wider to deal with traffic.
- Ability to connect multiple hardware or software entities, such as servers, so that they work as a single logical unit.
- This kind of scale cannot be implemented at a moment's notice.

SCALE DOWN / IN

Scaling DOWN : Vertical Scaling

- Removing resources from a single (powerful) node of the system.

Scaling IN : Horizontal Scaling

- Removing multiple (lesser capable) nodes from a systems.

LOCK-IN

- Vendor Lock-in / Proprietary lock-in or Customer lock-in
- Vendor lock-in is a situation in which customer using a product or service cannot easily transition to a competitor's product or service.
- Makes a customer dependent on a vendor for products & services, unable to use another vendor without substantial switching costs.
- Describes a service delivery technique that ensures customer dependence on the vendor services.
- Supplier may try to lock customer into the service.
  - SIM lock (network lock): A SIM card is proprietary to a specified phone manufacturer.

CLOUD LOCK-IN TYPES

- Platform Lock-in
- Data Lock-in
- Tool Lock-in

- There is no answer to eliminate cloud lock-in as such.
- Proper background study, planning & wise decision-making is the only reliable solution, as of now.