Cloud Architectural Concepts

What is Cloud Computing

- **NIST 800 - 145**
  - Cloud computing is a model for enabling **ubiquitous**, **convenient**, **on-demand** network access to a **shared pool** of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be **rapidly provisioned and released** with **minimal management** effort or service provider interaction.
  - This cloud model is composed of **five essential characteristics**, **three service models**, and **four deployment models**

Essential Characteristics of Cloud

1. Broad Network accessibility
2. On-demand Self Service
3. Rapid elasticity and scalability
4. Metered / Measured service
5. Resource Pooling

1. Broad Network Accessibility

1. Making the Compute available to any device from any where
2. Network connectivity, device compatibility and availability of compute environment are the focus

2. On-demand Self Service

1. Ability to subscribe to any compute or cloud service at anytime by any user without any dependency on anyone
2. A self-service mechanism to enable frictionless consumption of any service hosted in the cloud

2.1. On-demand Self Service

1. This self-service brings in a concept of **Shadow IT**
    1. use of information technology systems, devices, software, applications, and services without explicit IT department approval

Risks:

1. Loss of data and process control
2. Increased spending on un-authorized instances
3. Asset proliferation thereby increasing cost of operations and the attack surface

3. Rapid Elasticity and Scalability

1. Mechanism of increasing / decreasing the compute to the atomic value as required at anytime
2. This elasticity can be achieved through automatic or manual means

3.1 Elasticity and Scalability

- **Scalability**
  - Ability to grow as demand increases
  - Can be manual or automatic process
  - Scaling out and scaling up
- **Elasticity**
  - Ability to dynamically grow or shrink based on the resource utilization
  - Key feature that supports the economics of Cloud computing

All Elastic components are scalable, not all scalable components are Elastic

3.2 Elasticity and Scalability

Scaling Up	Scaling Out
1. Also called Vertical Scaling, its adding additional compute resources to the existing physical server	1. Also called Horizontal Scaling, its adding additional physical servers to supplement the existing server
2. There is tight coupling of the resources and hence performance is better	2. There is loose coupling of the resources and hence there is slight impact on the performance
3. Its simple and straight forward	3. Integrating multiple physical compute increases complexity
4. Powerful hardware at relatively lower energy consumption levels	4. As multiple small infra are added, there will be increase in power consumption. ,
5. There is a limit to Vertical scaling. It's the hardware limit beyond which you cannot scale up the server	5. There is theoretically no limit to the expansion
6. Higher configuration servers will result in higher cost of ownership	6. Since small infra can be added, the cost of ownership will not be high
7. Best suited for short term fixed limit expansion	7. Suited for long term increase in capacity

4. Metered / Measured Service

1. Financial accounting of consumption
2. Pay as you use model of consumption model
3. Helps the customer get a real economic operational expense for each of the business process IT infrastructure use

5. Resource Pooling

1. Virtualization of Hardware compute resources to monetize their use for multiple customers
2. Provides the ability to oversubscribe the available physical resources
3. Provides the economy of cloud computing

5.1. Resource Pooling

1. Resource pooling is the core of Multi-tenancy

2. It brings in few concepts on resource management

   1. **Reservations - Guarantee** a specific amount of resources to the VM

   2. **Limits** - Configure a **maximum** of what the VM can consume

   3. **Share** - Shares can be a **relative priority** when there is a resource contention. Shares come into play only when there is contention

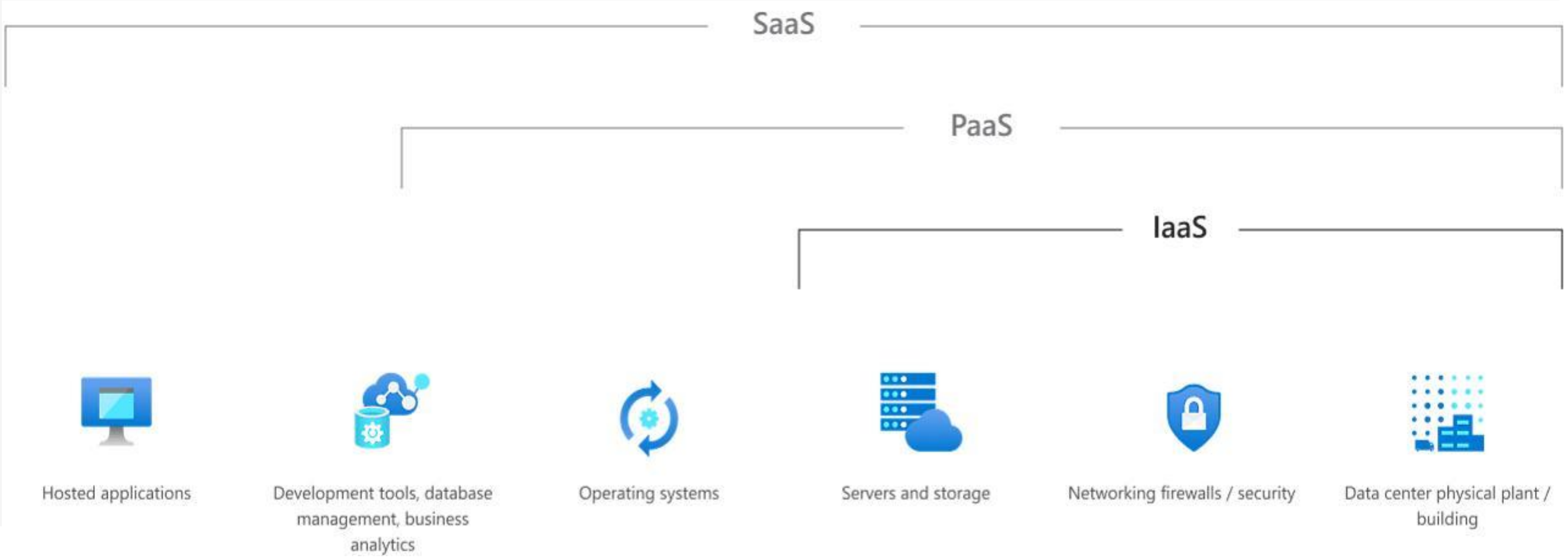
   4. **Borrow** - One VM can **temporarily borrow resources** from other VM pools. A runtime conflict can be triggered when the borrowed resource is not returned due to prolonged usage by the cloud service consumer that is borrowing it.

3 Service Models / Categories

1. Infrastructure as a Service (IaaS)

2. Platform as a Service (PaaS)

3. Software as a Service (SaaS)


SaaS
PaaS
IaaS
Hosted applications
Development tools, database management, business analytics
Operating systems
Servers and storage
Networking firewalls / security
Data center physical plant / building

Service Models - IaaS

- Cloud computing that delivers fundamental compute, network, and storage resources to consumers on-demand, over the internet, and on a pay-as-you-go basis

- IaaS provides the consumer Highest-level control of resources in the cloud

- Will have the lowest financial costs upfront, but the staff / Maintenance costs will be the highest

Service Models - PaaS

- Cloud computing model that provides the complete platform – HW / SW / App Framework/ OS / DB / Programming Libraries etc for the developer to develop and run their apps.

- This is the best model for customers who want to quickly and effectively test their applications across different flavors (provides a **runtime environment**)

Service Models - SaaS

- Cloud computing model where the application is hosted as a service by the provider and the customers subscribe to the service.
- Customer owns the data hosted in the SaaS application, rest of the stack is owned and managed by the Cloud Provider.
- It will typically have the highest startup and licensing costs, lowest costs in support staff and maintenance

4 Deployment Models

1. Public Cloud
2. Private Cloud
3. Community Cloud
4. Hybrid Cloud

Deployment Model – Public Cloud

- IT model where public cloud service providers make computing services available on-demand to organizations over the public internet

- They pool resources in distributed data centers around the world that multiple companies and users can access from the internet

- Public cloud helps companies to harness cutting-edge technologies and achieve global scale without shouldering the costs and labor themselves

- Reliability concerns can happen in public cloud – can be addressed by SLA

- A frequent risk is the retirement or removal of certain services that a customer relies on

Deployment Model – Private Cloud

- Single-Tenant environment setup on-premise or hosted by a third-party
- In this model, all hardware and software resources are dedicated exclusively to, and accessible only by, a **single customer**
- Most customizable, offers better access control, security and privacy
- Customer is granted exclusive access to the infrastructure
- In some cases, the customer may own the hardware
- Most costlier cloud deployment model, impedes elasticity/scaling

Deployment Model – Community Cloud

- Cloud infrastructure in which multiple organizations share resources and services based on common operational and regulatory requirements.
- Members of a community cloud are organizations that have common business requirements
- **Resiliency through shared ownership, shared costs**
- More expensive than the public cloud

Deployment Model – Hybrid Cloud

- Hybrid cloud refers to a **combination of at least 2 computing environments** that share information with one another and run a uniform series of applications for a business or enterprise
  - At least 1 private cloud and at least 1 public cloud
  - One bare-metal (physical hardware) or virtual environment connected to at least 1 cloud

Cost Benefit of Cloud Adoption

Reduction in Administrative personnel overhead

Reduction in Operational cost

Transferring some regulatory cost

Reduction in cost of Data Archival / Backup Services

Cloud Computing Roles

- Cloud Computing activities are grouped into 3 main groups:
  - activities that **use services**,
  - activities that **provide services** and
  - activities that **support services**

Cloud Roles

Cloud Provider

Cloud Customer / Consumer

Cloud Partner

Regulator

Cloud Broker

Cloud Service Customer

Cloud Service User

- Use Cloud Service

Cloud Service Administrator

- Perform service trail
- Monitor service
- Administer security service
- Provide billing and usage reports
- Handle problem reports
- Administer Tenancies

Cloud Service Business Manager

- Perform business administration
- Select and purchase service
- Request audit report

Cloud Service Integrator

- Connect on-premise systems to cloud

Customer is ultimately responsible for the data

Cloud Service Provider

- Company or other entity offering cloud services
- They make the services available
  - Service Deployment
  - Service Orchestration
  - Service Management
- They are responsible for the Physical and Environmental Infrastructure of the Cloud Data Centres.
- They take appropriate measures to identify the service capacity they need to provision
- In a Multi-tenancy environment providing Isolation is the MOST crucial security task of the CSP

Cloud Service Provider

Cloud Service Manager

- Prepare the systems
- Provide services
- Manage business plan
- Monitor and administer services
- Provide audit data
- Perform SLA mgmt.
- Manage customer relationship

Customer Support Engineer

- Handle customer request

Security and Risk Manager

- Manage security and Risks
- Design and implement service continuity
- Ensure compliance

Network Provider

- Manage peer cloud services
- Provide network connectivity
- Perform federation and integration services

Cloud Service Partner

Cloud Service Partner

- Design, create and maintain service components
- Compose services
- Test the services

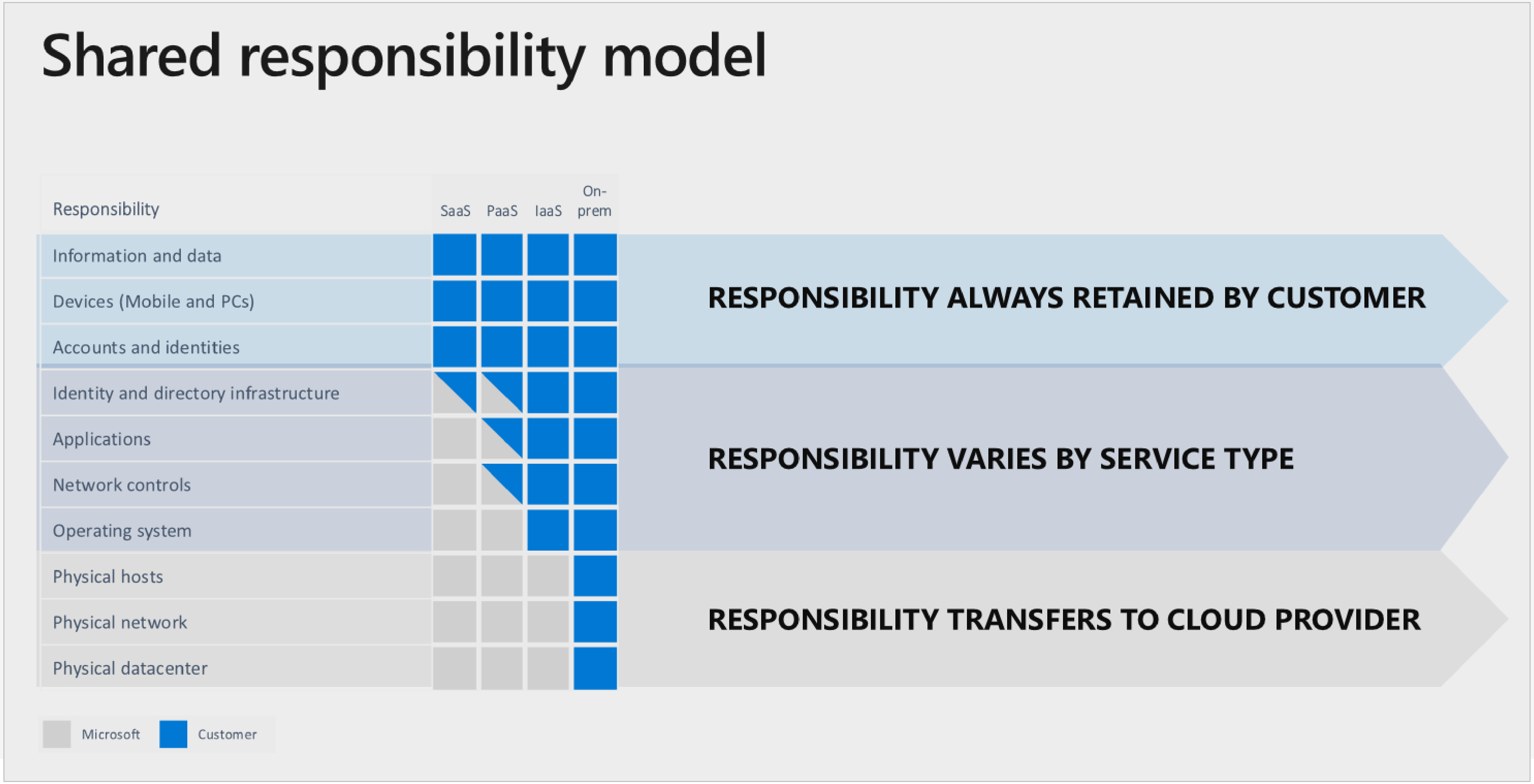
Auditor

- Perform Audit
- Report audit results

Cloud Service Broker

- Acquire and assess the customers
- Assess marketplace
- Setup legal agreement
- Package services
- Aggregate, integrate and customize the cloud services

Shared Responsibility Model



Shared responsibility model

Responsibility	SaaS	PaaS	IaaS	On-prem
Information and data	🟦	🟦	🟦	🟦
Devices (Mobile and PCs)	🟦	🟦	🟦	🟦
Accounts and identities	🟦	🟦	🟦	🟦
Identity and directory infrastructure	◥	◥	🟦	🟦
Applications	⬜	◥	🟦	🟦
Network controls	⬜	◥	🟦	🟦
Operating system	⬜	⬜	🟦	🟦
Physical hosts	⬜	⬜	⬜	🟦
Physical network	⬜	⬜	⬜	🟦
Physical datacenter	⬜	⬜	⬜	🟦

RESPONSIBILITY ALWAYS RETAINED BY CUSTOMER

RESPONSIBILITY VARIES BY SERVICE TYPE

RESPONSIBILITY TRANSFERS TO CLOUD PROVIDER

⬜ Microsoft 🟦 Customer

Cloud Building block Technologies

What is Virtualization

- Virtualization is a process that allows for more efficient utilization of physical computer hardware
- It is the foundation of cloud computing
- Virtualization uses software to create an abstraction layer over computer hardware that allows the hardware elements of a single compute to be divided into multiple virtual computers (VMs)
- Each VM runs its own operating system (OS) and behaves like an independent compute

Components

- **Virtual machines (VMs):**
  - Virtual machines (VMs) are virtual environments that simulate a physical compute in software form.
- **Hypervisors**:
  - Software layer that coordinates VMs
  - It serves as an interface between the VM and the underlying physical hardware
  - It also ensures that the VMs don’t interfere with each other

Hypervisor

- **Type 1 Hypervisor**
  - “bare-metal” hypervisors interact with the underlying physical resources
  - Is faster and more secure
  - Most common Virtualization found in Data Centers
- **Type 2 Hypervisor**
  - Hypervisors run as an application on an existing OS
  - Most commonly used on endpoint devices
  - They carry a performance overhead because they must use the host OS
  - A Type 2 hypervisor on a host OS is also susceptible to host OS compromise

Threats

- **Virtual Machine Escape:**
  - VM escaping is a vulnerability in virtualization technology where an attacker escapes the isolation of a virtual machine and gains access to the underlying operating system and other VMs on the **same physical machine**
- **Host Escape**
  - Host escape is a vulnerability in virtualization where an attacker escapes the Hypervisor and can gain access to adjacent machines in the network.

Containers

Containers

- A container is a standard unit of software that packages up code and all its dependencies, so the application runs quickly and reliably from one computing environment to another
- In Containerization, there is no hypervisor or guest OS
- A container runtime sits above the host OS, and each container uses the runtime to access the needed system resources
- Increased portability and light weight are the key advantages of containers

Cryptography and Key Management

Cryptography and Key Management

- Cryptography is fundamental to cloud Computing
- 3 Key challenges in cloud that requires Cryptography
  - Data in transit protection
  - Multitenancy and
  - the inability to physically swipe the hard disks

Key Management

- Main challenge in Cryptography is Key Management
- **Key Management Service (KMS)**
  - Stores keys separately from the data
  - Monitoring, automation and auditing are key questions when evaluating a KMS
- **Hardware Security Modules (HSM)**
  - Physical devices that provide key management services

Media Sanitization

Sanitization

- **Clearing (overwriting/wiping/shredding)**
  - Process of preparing media for reuse with assurance that cleared data cannot be retrieved using traditional recovery means
  - Unclassified data is written over all addressable locations on the media
  - Data recovery requires special laboratory techniques
  - Some media types don’t respond well to clearing
- **Purging**
  - More intense form of clearing – repeats the clearing process multiple times
  - Provides assurance that data cannot be recovered using any known means
  - It can be combined with other means like degaussing to completely remove data

Sanitization

- **Declassification**
  - Any process that purges media or system for reuse in unclassified environment
- **Sanitization**
  - Combination of process that ensures data is removed from the system
  - It ensures data cannot be recovered by any means
  - Includes ensuring non-volatile memory is erased, external drives removed and sanitized
- **Degaussing**
  - Generates heavy magnetic fields which realign the magnetic fields in magnetic media, only effective on magnetic media (does not affect, CD/DVD/SSD)
- **Cryptographic Erasure:**
  - The data to be protected is encrypted using strong cryptographic keys
  - Post encryption process, the keys are securely destroyed eliminating the key necessary for decryption

Other Terminologies

Edge and Fog Computing

- **Fog Computing**
    - **Decentralized infrastructure** that **places storage and processing components at the edge of the cloud**, where data sources such as application users and sensors exist
    - Fog computing **involves the usage of devices with lower processing capabilities to share some of the cloud's load**
    - The goal of fog computing is to use the cloud only for long-term and resource-intensive analytics, while the edge devices take care of short-term and time-critical analytics such as fault alerts, alarm status, etc
- **Edge Computing:**
    - Distributed computing framework that brings enterprise applications closer to data sources such as IoT devices or local edge servers.
    - Edge computing is a **subset of fog computing** that involves **processing data right at the point of creation**

Confidential Computing

- Confidential computing is a cloud computing technology that isolates sensitive data in a protected CPU enclave during processing

- The contents of the enclave are accessible only to authorized programming code, and are invisible and unknowable to anything or anyone else, including the cloud provider

- This is achieved by offering **Trusted Execution environments (TEEs)**

Ephemeral Computing

- Ephemeral computing refers to resources that are created when needed and immediately deprovisioned when no longer needed
- Traditional monitoring models do not work for this transient nature of these devices
- The key to secure Ephemeral computing is to properly specifying configuration of the ephemeral asset
- Infrastructure as Code (**IaaC**) is a very good model for managing Ephemeral computing assets

Serverless Computing

- CSP is solely responsible for maintaining the servers and exposes their computing capacity to customers
- When serverless code is run, the CSPs serverless environment allocates resources, the results are stored in persistent memory and the requesting function reads the data from this memory store
- Serverless applications rely on APIs for communication and executing Functions

Baselining

- Refers to known set of configuration attributes for a system
- It is important to ensure new systems / applications follow the defined baseline
- ***Immutable architecture*** prohibits changes to environments once they are built ~ it is important to baselines

Machine Learning

- ML applies the principles of Data science and statistics to uncover knowledge hidden in the data
- **Descriptive Analytics**: seeks to describe the data
- **Predictive Analytics**: seeks to use existing data to predict the future event
- **Prescriptive Analytics**: seek to **optimize** our behavior by simulating many scenarios

Well-Architected Framework

- Best Practices that can be used to evaluate and manage cloud infrastructure
- Key offerings are:
  - Security
  - Reliability
  - Performance

ISO Standards

- **ISO 27017** – Guides implementation of controls in the cloud computing environment
- **ISO 27018** – Guides the implementation of controls for **protection of PII processed** in the cloud; it is a supplement of ISO 27002

Government Cloud Standards

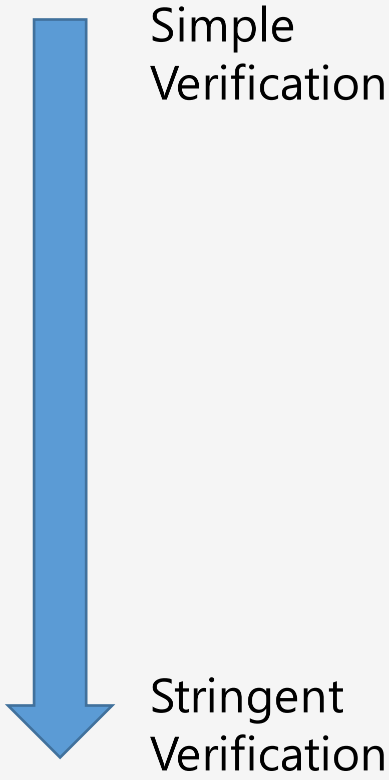
FedRAMP	Provides a set of risk-based security controls CSP has to implement. An audit process is designed to validate the implementation.
UK G-Cloud	Marketplace for Cloud services that have implemented controls in line with the G-Cloud Framework
CSA STAR	Voluntary scheme with two levels of assurance: **Level 1** : Self-assessment, CSP completes a CSA CCM and submits it to the STAR Registry **Level 2:** CSP undergoes Third-party audit **Level 3**: CSP is under continuous monitoring by Third-party
Common Criteria	Set of guidelines and specifications to evaluate information security **products**. There are two parts: **Protection Profile**: Defines the standard set of security requirements for a specific product type **Evaluation Assurance Level**: Scores between 1 to 7, 7 being the highest level.
FIPS 140-3	Provides scheme for validating the strength of cryptographic modules

CC

- Evaluates the products against protection profile
- Evaluated products are assigned Evaluation Assurance Levels [EAL]
- Addresses Functionality and Assurance
- ISO/IEC 15408 is used as the basis for evaluation of security properties
  - 15408-1: Introduction and general evaluation model
  - 15408-2: Security functional components
  - 15408-3: Security Assurance components

CC

- EAL1: Functionally Tested
- EAL2: Structurally Tested
- EAL3: Methodically tested and checked
- EAL4: Methodically designed, tested and checked
- EAL5: Semi formally designed and tested
- EAL6: Semi formally verified design and tested
- EAL7: Formally verified design and tested



Simple
Verification

Stringent
Verification

Key definitions

Terminologies

Cloud Bursting	Cloud bursting is a configuration method that uses cloud computing resources whenever on-premises infrastructure reaches peak capacity. Cloud bursting is a convenient and cost-effective way to to support workloads with varying demand patterns and seasonal spikes in demand.
Multitenancy	Multitenancy is when several different cloud customers are accessing the same computing resources Multitenant architecture is a feature in many types of public cloud computing
Oversubscription	The oversubscription of resources in cloud computing happens when a shared hosting or Public Cloud provider offers a series of computing resources that exceed the available capacity
Reversibility	Reversibility is the extent to which cloud-based applications are designed so that they can be moved to other cloud providers or on-premise environments.
Vendor Lock-in	Customer is forced to continue using a product or service regardless of quality, because switching away from that product or service is not practical
Vendor Lock-out	Customer faces negative impact on business due to the cloud vendor not providing the services and the customer is not able to move out due to dependency issues

Terminologies

Interoperability	• The ability of the systems to work efficiently and collaborate effectively across different cloud platforms • The capability of moving an application from one cloud service to another or between a client's environment and a cloud service • It is the ease with which one can move or reuse components of an application or service • The main concept is to not have such dependencies on the underlying operating system, hosting environment, libraries, or APIs that lock in a service to one particular set of hosts or solutions
Portability	• Cloud portability is the ability of a cloud computing product, solution or service to be migrated to a new vendor or location without incurring substantial porting and integration issues. • For portability, the **main focus is not on reuse or repurposing**, but rather solely on the ability to freely and easily move.
Provisioning	• The process by which credentials are verified and proven, and then an authentication object or token is granted for access.
Authorization	• The process by which appropriate roles and permissions are granted to the user or service account

Business Continuity

Business Continuity

- Used to maintain the continuous operations of business-critical functions in the event of a disaster
- **Disaster Recovery**
  - Goal is to minimize the immediate effects of a disaster.
  - IT focused
- **Continuity Planning**
  - Provides methods and procedures for long term outages and disasters
  - It takes a broader approach to disaster recovery
- **4 Main process steps for BCP**
  1. Project scope and planning
  2. Business Impact assessment (BIA)
  3. Continuity planning
  4. Approval and Implementation

1. Project scope and planning

- First step in effective BCP
- Involves 4 key work streams
  - Structured analysis of business organization from **crisis point of view**
  - **Creation of BCP team** with approval from senior management
  - **Resource availability** assessment
  - Legal and Regulatory requirement analysis

Business Organization Analysis

- First step is to perform an analysis of business organization to identify all critical departments and key stakeholders
    - Operational departments that are responsible for core services
    - Critical support services, responsible for upkeep of systems that support operational departments
    - Senior executives and other key Individuals essential for ongoing business operations
- This step provides necessary groundwork to identify potential members of BCP team
- It provides **foundation for the business continuity process**

BCP Team Selection

- The team should include at the minimum the following representatives
  - Representatives from each of the organizational departments
  - Representatives from key support departments
  - IT representatives with technical expertise
  - Security representatives with knowledge of BCP
  - Legal representatives
  - Senior Management representatives

BCP Resource Requirement

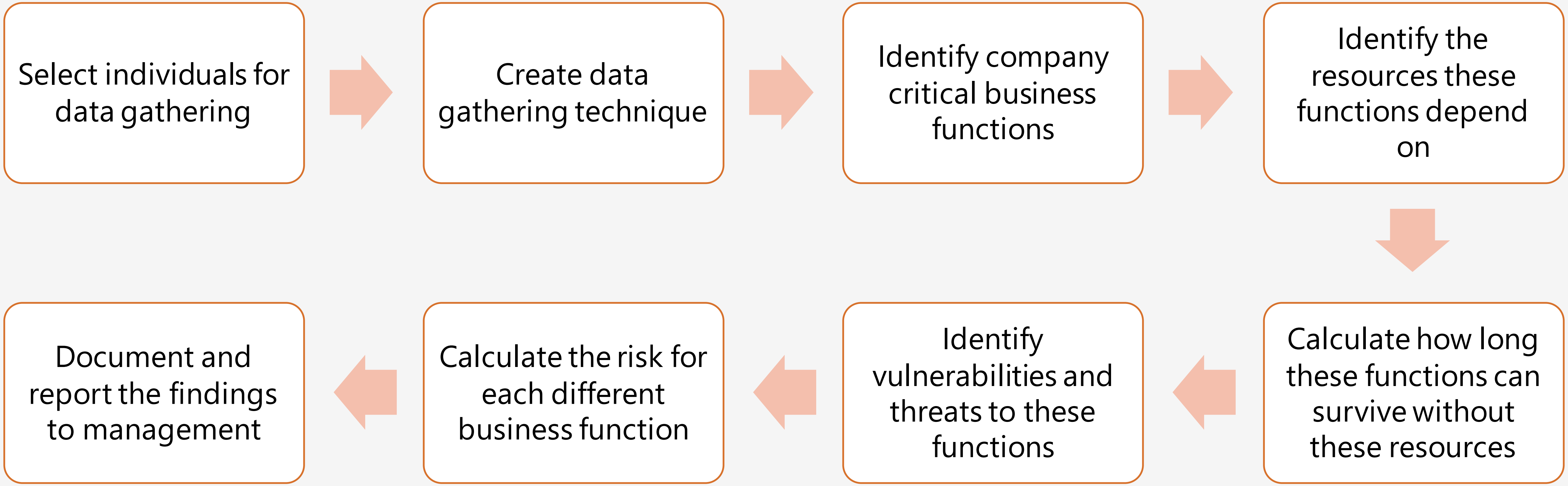
- Assess the resource required for 3 distinct functions
- **BCP Development:**
  - Team will require resources for BCP process development.
- **BCP Testing, training and maintenance:**
  - Will require hardware and software commitments, major commitment will be the people
- **BCP Implementation:**
  - Implementation will require a large number of resources both from the HW/SW as well as human capital front

Human capital is the most significant resources consumed during a BCP process

Business Impact Assessment

- Also considered a ***functional analysis***
- Identifies the resources that are critical to the organization, the threats posed to the resources
- Assess the likelihood that each threat will actually occur and the impact of those threats
- The result helps in prioritizing the commitment of Business continuity resources to various risk exposures

Business Impact Assessment - Steps



Select individuals for data gathering

Create data gathering technique

Identify company critical business functions

Identify the resources these functions depend on

Calculate how long these functions can survive without these resources

Identify vulnerabilities and threats to these functions

Calculate the risk for each different business function

Document and report the findings to management

1. Identifying Priorities

- 1st step in BIA is to identify the business priorities
- It involves creating a comprehensive list of business process and ranking them in order of importance
- This is a qualitative process; to begin quantitative assessment, assign AV in monetary terms to each asset
- Develop the **Maximum Tolerable Downtime (MTD)**
  - Maximum time the business can be inoperable without causing irrecoverable damage to the business
- Develop the **Recovery Time Objective (RTO)**
  - Amount of time by which the business function can be recovered

GOAL – RTO must be less than MTD

2. Risk Identification

- Next step in BIA process
- Risk comes in two forms : Man-made or Natural
- The risk identification portion of the process can be qualitative or quantitative
- BCP team should not be concerned about likelihood or the amount of damage in this phase

3. Likelihood Assessment

- Follows the Risk Identification Phase
- Identifies the likelihood that each risk will occur
- It is expressed in ARO
- ARO should be based on company history, professional experience of team members and advice from experts

4. Impact Assessment

- Most critical portion of BCP
- Analyse the data gathered during risk identification and Likelihood assessment to determine what impact each one of the identified risks would have on the business

Continuity Planning

- Next Phase of BCP process, focuses on developing and implementing a continuity strategy to minimize the impact realized-risks might have on protected assets
- **Strategy Development**:
  - Bridges the gap between BIA and continuity planning phases.
  - Take the risks identified and determine which risks will be addressed by BCP

Continuity Planning Goals

- To ensure continuous operations of business in the face of an emergency
- **Statement of Importance** is a letter to the Organization stakeholders from the CEO/Board stating the efforts being taken in developing a effective BCP Strategy
- **Statement of Priorities** involves listing the functions identified as critical for continuous operations.

Continuity Planning Goals

- **Statement of Organizational Responsibility** states organizations commitment to business continuity planning.
- **Statement of Urgency and Timing** expresses the criticality of implementing the BCP and outlines the timetable

BCP Policy

- BCP policy benefits
  - Ensures BCP professionals have a written continuity document to reference in the event of an emergency
  - Provides historical record of the BCP that will be useful to future personnel
  - Forces the team members to commit their thoughts to paper

Cloud-Specific BIA concerns

- **New Dependencies**
  - Reliance of third-parties and their downstream entities for operations
- **Regulatory Failure**
  - Dispersion can bring in potential regulatory violations
  - Provider may be unable to comply to mandated regulations
  - Weak Contractual clauses
- **Data Breach / Inadvertent Disclosure**
  - Magnification in the likelihood and impact of risks associated with cloud
  - Public disclosure of internal communication, loss of competitive advantage, negative effect on customer, supplier and vendor goodwill
- **Vendor Lock-In / Lock-Out**

Cloud-Specific BC/DR Responsibilities

- **Logical Location of Backup Data / Systems**
  - 3 general means of using cloud backups for BC / DR
    - **Private Architecture, Cloud Service as Backup**
    - **Cloud Operations, Cloud Provider as Backup**
    - **Cloud Operations, Third-Party Cloud Backup Provider**

Private Architecture, Cloud Service as Backup

- Organization maintains its own DC for primary production environment, uses cloud for BC / DR purpose
  - Negotiations should include, upload bandwidth costs, frequency of backups, type of backups, security of the data and systems and the ISP costs.
  - Customer will determine when the failover will occur and when normal operations will cease and the backup will be utilized
  - Customer will issue formal notification to the provider
  - Failover can take any form (IaaS, PaaS, SaaS) or just download of the backup data to another location
  - It should include how and when the download will happen, how long it should take and how and when the data will be restored to normal operations

Cloud Operations, Cloud Provider as Backup

- Organization uses a cloud provider for production environment, and also subscribe to backup operations from the same provider
    - Backups will be stored in a geographically different location to provide resiliency
    - Provider will be responsible for determining the location, backup configuration
    - Customer will have minimal participation in the failover process ~ transparent to the user
    - Normally provided at no or little cost addition
    - Typically SaaS services are this model

Cloud Operations, Third-Party Cloud Backup Provider

- Organization uses a cloud provider for production environment, subscribed to another cloud provider for Backup
  - Customers who want to distribute risk, enhance redundancy, prevent Vendor Lock-In / Lock-Out prefer this model
  - Most complicated model involving preparations and coordination between multiple parties
  - Emergency assessment, declaration and failover will be a joint effort among customer, primary and secondary cloud provider
  - Potentially expensive model
  - Can bring in portability and interoperability concerns

All the best