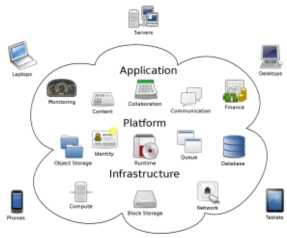


UNIT III

Cloud Computing

Contents

UNIT-3

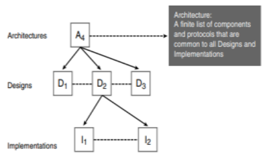
- Data Centre Architecture and Technologies
  - Architectural Building Blocks of Data Centre
  - Industry Direction and Overview of Operational and Technical Phasing
    - Overview of 5 Phases
- Computing with Titans
  - Google
  - Microsoft
  - Amazon
  - IBM
  - Accessing the Cloud- Platforms through
    - Web Applications
    - Web API's
    - Web Browsers

Architecture

- Architects are typically individuals who have extensive experience in building IT systems that meet specific business requirements and translating those business requirements into IT engineering requirements.
- subject matter experts to interpret the high-level architectural requirements into a low-level design and ultimately implement (build) a system ready for use
- Architecture: A finite list of components and protocols that are common to all designs and implementation
- With regard to cloud services, architecture must extend beyond on-premises (private cloud) deployments to support hybrid cloud models (hosted cloud, public cloud, community cloud, virtual private cloud, and so on).
- Architectural principles that are required for a services platform today would most likely include but not be limited to efficiency, scalability, reliability, interoperability, flexibility, robustness, and modularity.
- To implementing and managing architecture, process frameworks and methodologies are now heavily utilized to ensure quality and timely delivery by capitalizing of perceived industry best practices

Architecture

- Many-to-one relationship among architecture, design, and implementations
- A Well-understood communication among all stakeholders is essential throughout the project delivery phases to ensure success



Data Centre Architecture and Technologies

- Data center design is at an evolution for:
  - Massive data growth
  - challenging economic conditions
  - the physical limitations of power
  - Heat
  - space
- how they are supporting architectural principles outlined previously?
- how they are influencing design and implementation of infrastructure?
- Architectural, Design and Implementation value in regard to delivering IT as a service

Data Centre Architecture and Technologies

- Data center IT staff is typically asked to address the following data center challenges
  - Improve asset utilization to reduce or defer capital expenses
  - Reduce capital expenses through better management of peak workloads
  - Make data and resources available in real time to provide flexibility and alignment with current and future business agility needs.
  - Reduce power and cooling consumption to cut operational costs and align with "green" business practices.
  - Reduce deployment/churn time for new/existing services, saving operational costs and gaining competitive advantage in the market.
  - Enable/increase innovation through new consumption models and the adoption of new abstraction layers in the architecture.
  - Improve availability of services to avoid or reduce the business impact of unplanned outages or failures of service components.
  - Maintain information assurance through consistent and robust security posture and processes.

Technology to Support Architectural Principles

- Efficiency
  - Virtualization of infrastructure with appropriate management tools
  - Infrastructure homogeneity is driving asset utilization up
- Scalability
  - Platform scalability can be achieved through explicit protocol choice and hardware selection and also through implicit system design and implementation
- Reliability
  - Disaster recovery (BCP) planning, testing, and operational tools (for example, VMware's Site Recovery Manager, SNAP, or Clone backup capabilities)
- Interoperability
  - Web-based (XML) APIs, for example, WSDL (W3C) using SOAP or the conceptually simpler RESTful protocol with standards compliance semantics, for example, RFC 4741 NETCONF or TMForum's Multi-Technology Operations Systems Interface (MTOSI) with message binding to "concrete" endpoint protocols
- Flexibility
  - Software abstraction to enable policy-based management of the underlying infrastructure. Use of "meta models" (frames, rules, and constraints of how to build infrastructure). Encourage independence rather than interdependence among functional components of the platform

Technology to Support Architectural Principles

- Modularity
  - Commonality of the underlying building blocks that can support scale-out and scale-up heterogeneous workload requirements with common integration points (web-based APIs). That is, integrated compute stacks or infrastructure packages (for example, a Vblock or a FlexPod). Programmatic workflows versus script-based workflows along with the aforementioned software abstraction help deliver modularity of software tools.
- Security
  - The appropriate countermeasures (tools, systems, processes, and protocols) relative to risk assessment derived from the threat model. Technology countermeasures are systems based, security in depth. Bespoke implementations/design patterns required to meet varied hosted tenant visibility and control requirements necessitated by regulatory compliance.
- Robustness
  - System design and implementation—tools, methods, processes, and people that assist to mitigate collateral damage of a failure or failures internal to the administratively controlled system or even to external service dependencies to ensure service continuity

Industry Direction and Operational and Technical Phasing

- New technologies, such as multicore CPU, multi-socket motherboards, inexpensive memory, and Peripheral Component Interconnect (PCI) bus technology, represent an evolution in the computing environment
- The advancements in technologies leads to access the greater performance and resource utilization at a time of exponential growth of digital data and globalization via Internet.
- bandwidth intensive and require higher performance and efficiency from the underlying infrastructure used by multithreaded application
- Basic Hypervisor Technology:
  - virtual switches embedded in the hypervisor/VMM kernel to bring together operational domain of virtual servers and the network
  - delivering consistent and integrated policy deployments
- To Live migration of a VM
  - require orchestration of (physical and virtual) server, network, storage, and other dependencies to enable uninterrupted service continuity

Industry Direction and Operational and Technical Phasing

- Data center performance requirements are growing
- IT managers are seeking ways to limit physical expansion by increasing the utilization of current resources.
  - Server consolidation by means of server virtualization has become an appealing option.
  - The use of multiple virtual machines takes full advantage of a physical server's computing potential and enables a rapid response to shifting data center demands
  - Increase in computing power, coupled with the increased use of VM environments, is increasing the demand for higher bandwidth and at the same time creating additional challenges for the supporting networks.
- Power consumption and efficiency
- Metrics to determine the efficiency of Data Center
  - Facilities
  - Network
  - Server
  - Storage systems
  - the availability, power capacity, and density profile of mail, file, and print services will be very different from those of mission-critical web and security services

Industry Direction and Operational and Technical Phasing

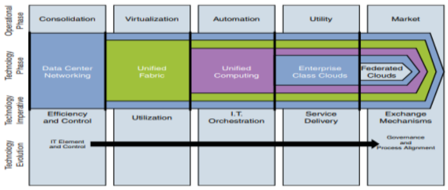
- Current Barriers to Cloud/Utility Computing
  - Lack of trust in current cloud offerings is the main barrier to broader adoption of cloud computing
  - Without trust, the economics and increased flexibility of cloud computing make little difference
    - For example, from a workload placement perspective, how does a customer make a cost-versus-risk (Governance, Risk, Compliance [GRC]) assessment without transparency of the information being provided
  - Transparency requires well-defined notations of service definition, audit, and accountancy
  - Multiple industry surveys attest to Transparency in Cloud Computing

Industry Direction and Operational and Technical Phasing

- Challenges to Trust in Cloud
  - Security
    - sufficient information assurance (IA) processes and tools to enforce
      - confidentiality, integrity, and availability of the corporate data assets
  - Control
    - control to decide how and where data and software are deployed, used, and destroyed in a multitenant and virtual, morphing infrastructure
  - Service-level management
    - appropriate Resource Usage Records (RUR) be obtained and measured appropriately for accurate billing
    - each application get the necessary resources and priority needed to run predictably in the cloud
  - Compliance:
    - cloud environment conform with mandated regulatory, legal, and general industry requirements
  - Interoperability
    - vendor lock-in given the proprietary nature of today's public clouds
    - ability to reduce risk through "multi-homing" network connectivity to multiple Internet service providers

Operational and Technological Evolution Stages of IT

- Solve Fundamental challenges in the data center
- The direction in which the IT Industry is heading

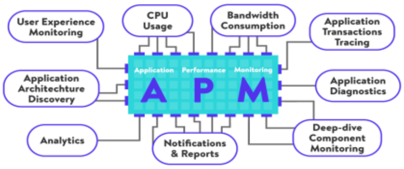


Industry Direction and Operational and Technical Phasing

- **Adoption of a broad IP WAN that is highly**
  - enables centralization and consolidation of IT services, application-aware services are layered on top of the WAN to intelligently manage application performance
- **Executing on a virtualization strategy for server, storage, networking, and networking services**
  - enables greater flexibility in the substantiation of services in regard to physical location, thereby enabling the ability to arrange such service to optimize infrastructure utilization.
- **Service automation**
  - enables greater operational efficiencies related to change control, ultimately paving the way to an economically viable on-demand service consumption model
- **Utility computing model**
  - includes the ability meter, chargeback, and bill customer on a pay-as-you-use (PAYU) basis
- **Market**
  - creation through a common framework incorporating governance with a service ontology that facilitates the act of arbitrating between different service offerings and service providers.

Phase 1: The Adoption of a Broad IP WAN That Is Highly Available

- Many applications were written to operate over a LAN and not a WAN environment
- The optimal economic path forward is to utilize application-aware, network-deployed services to enable a consistent Quality of Experience (QoE) to the end consumer of the service without rewriting the code.
- Application Performance Management (APM)
  - includes capabilities such as visibility into application response times, analysis of which applications and branch offices use how much bandwidth, and the ability to prioritize mission-critical applications



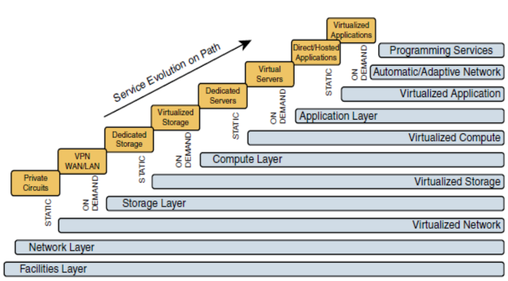
Phase 1: The Adoption of a Broad IP WAN That Is Highly Available

- Specific capabilities to deliver APM are as follows:
  - Performance monitoring
    - Both in the network (transactions) and in the data center (application processing).
  - Reporting:
    - Application SLA reporting requires service contextualization of monitoring data to understand the data in relation to its expected or requested performance parameters. These parameters are gleaned from who the service owner is and the terms of his service contract.
  - Application visibility and control:
    - Application control gives service providers dynamic and adaptive tools to monitor and assure application performance.

Phase 2: Executing on a Virtualization Strategy for Server, Storage, Networking, and Networking Services

- Virtualization is the concept of creating a "sandbox" environment, where the computer hardware is abstracted to an operating system
- VM
- Virtualization technology can also be applicable to many different areas such as networking and storage
- virtual LAN (VLAN) and routing with Virtual Routing and Forwarding (VRF) tables
- storage-area networks have something similar in terms of virtual storage-area networks (VSAN)
- vFiler for NFS storage virtualization
- price to pay for all this virtualization: management complexity
  - Both implicit within infrastructure components or explicitly in external management tools, are required to provide the visibility and control service operations teams required to manage the risk to the business
- virtual applications (vApp) have also been brought into the data center architecture to provide policy enforcement within the new virtual infrastructure, again to help manage risk

Phase 2: Executing on a Virtualization Strategy for Server, Storage, Networking, and Networking Services



IT Service Enablement Through Abstraction/Virtualization of IT Domains

Phase 3: Service Automation

- Service automation, working hand in hand with a virtualized infrastructure, is a key enabler in delivering dynamic services
- The policy-driven provisioning of IT services though the use of automated task workflow, whether that involves business tasks (also known as Business Process Operations Management [BPOM]) or IT tasks (also known as IT Orchestration).
- Too Costly, because of rewriting many scripts for automating the services
  - if an architect wants or needs to change an IT asset (for example, a server type/supplier) or change the workflow or process execution logic within a workflow step/node in response to a business need, a lot of new scripting is required
- Two main developments to made service automation a more economically:
  - Standards-based web APIs and protocols (for example, SOAP and RESTful) have helped reduce integration complexity and costs through the ability to reuse.
  - Programmatic-based workflow tools helped to decouple/abstract workflow from process execution logic from assets
    - Contemporary IT orchestration tools, such as Enterprise Orchestrator from Cisco and BMC's Atrium Orchestrator, allow system designers to make changes to the workflow (including invoking and managing parallel tasks) or to insert new workflow steps or change assets through reusable adaptors without having to start from scratch. Using the LEGO wall analogy, individual bricks of the wall can be relatively easily interchanged without having to build a new wall.

Phase 3: Service Automation

- Third component is necessary to make programmatic service automation a success, namely, an *intelligent infrastructure* by which the complexity of the low-level device configuration syntax is abstracted from the northbound system's management tools.
- The higher-level management tools only need to know the policy semantics
  - In other words, an orchestration system need only ask for a *chocolate cake* and the element manager, now based on a well-defined (programmatic) object-based data model, will translate that request into the required ingredients and, furthermore, how they those ingredients should be mixed together and in what quantities.
- A practical example is the Cisco Unified Compute System (UCS) with its single data model exposed through a single transactional-based rich XML API
  - UCS provides a layer of abstraction between its XML data model and the underlying hardware through application gateways that do the translation of the policy semantics as necessary to execute state change of a hardware component

Phase 4: Utility Computing Model

- the ability to
  - monitor, meter, and track resource usage for chargeback billing
- self-service provisioning (on-demand allocation of compute resources), in essence turning IT into a utility service
- it is crucial to maintain knowledge of allocation and utilization of resources
- Metering and performance analysis of the resources enable
  - cost efficiency, service consistency, and subsequently the capabilities IT needs for trending, capacity management, threshold management
- Pay for use, where the end customers are charged based on their usage and consumption of a service
- *pay-per-use* has gained acceptance in enterprise computing as IT works in parallel to lower costs across infrastructures, applications, and services
- IT providers have typically struggled with billing solution metrics that do not adequately represent all the resources consumed as part of a given service

Phase 4: Utility Computing Model

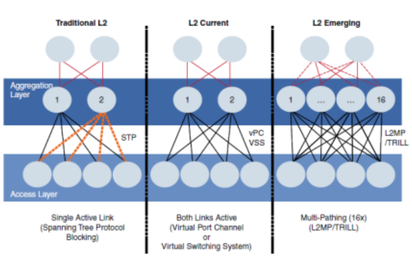
- The dynamic nature of a virtual converged infrastructure and its associated layers of abstraction being a benefit to the IT operation conversely increase the metering complexity
  - optimal chargeback solution provides businesses with the true allocation breakdown of costs and services delivered in a converged infrastructure
- The business goals for metering and chargeback typically include the following
  - Reporting on allocation and utilization of resources by business unit or customer
  - Developing an accurate cost-to-serve model, where utilization can be applied to each user
  - Providing a method for managing IT demand, facilitating capacity planning, forecasting, and budgeting
  - Reporting on relevant SLA performance
- Chargeback and billing requires three main steps:
  - **Step 1.** Data collection
  - **Step 2.** Chargeback mediation (correlating and aggregating data collected from the various system components into a billing record of the service owner customer)
  - **Step 3.** Billing and reporting (applying the pricing model to collected data) and generating a periodic billing report

Phase 5: Market

- In mainstream economics, the concept of a market is any structure that allows buyers and sellers to exchange any type of goods, services, and information
- The exchange of goods or services for money (an agreed-upon medium of exchange) is a transaction
- For a marketplace to be built to exchange IT services as an exchangeable commodity, the participants in that market need to agree on common service definitions or have an ontology that aligns not only technology but also business definitions
- The alignment of process and governance among the market participants is desirable
- The service components from different providers/authors to deliver an end-to-end service
- A service has two aspects
  - Business
    - required for marketplace and a technical aspect for exchange and delivery
    - product definition, relationships (ontology), collateral, pricing, and so on
  - Technical
    - fulfillment, assurance, and governance aspects

Design Evolution in the Data Center

- emerging technologies in the data center
  - how they are supporting architectural principles outlined previously
  - how they are influencing design and implementation of infrastructure, and ultimately their value in regard to delivering IT as a service



Design Evolution in the Data Center

- It was necessary to develop protocol and control mechanisms that limited the disastrous effects of a topology loop in the network
- STP provides a loop detection and loop management capability for Layer 2 Ethernet networks
- Extensions and enhancement of STP, scales to very large networks with one suboptimal principle
  - To break loops in a network, only one active path is allowed from one device to another
  - the single logical link creates two problems
    - Half (or more) of the available system bandwidth is off limits to data traffic
    - A failure of the active link tends to cause multiple seconds of system-wide data loss while the network reevaluates the new "best" solution for network forwarding in the Layer 2 network.
- Virtual PortChannel (vPC)
- Layer 2 Multi-Pathing (L2MP)

Design Evolution in the Data Center

- Virtual PortChannel (vPC)
  - In Large networks, the support of multiple devices together is often a design requirement to provide some form of hardware failure alternate path
  - This alternate path is often connected in a way that would cause a loop, limiting the benefits gained with PortChannel technology to a single path
  - To address this limitation, the Cisco NX-OS Software platform provides a technology called virtual PortChannel (vPC)
  - a pair of switches acting as a vPC peer endpoint looks like a single logical entity to PortChannel-attached devices, the two devices that act as the logical PortChannel endpoint are still two separate devices
  - STP can recover from a link failure in approximately 6 seconds, while an all-PortChannel-based solution has the potential for failure recovery in less than a second

Design Evolution in the Data Center

- Layer 2 Multi-Pathing (L2MP)
  - IETF Transparent Interconnection of Lots of Links (TRILL) is a new Layer 2 topology based capability. With the Nexus 7000 switch, Cisco already supports a prestandards version of TRILL called FabricPath, enabling customers to benefit from this technology
- The operational benefits of L2MP are as follows
  - Enables Layer 2 multipathing in the Layer 2 DC network (up to 16 links). This provides much greater cross-sectional bandwidth for both client-to-server (North-to- South) and server-to-server (West-to-East) traffic.
  - Provides built-in loop prevention and mitigation with no need to use the STP. This significantly reduces the operational risk associated with the day-to-day management and troubleshooting of a nontopology-based protocol, like STP.
  - Provides a single control plane for unknown unicast, unicast, broadcast, and multicast traffic.
  - Enhances mobility and virtualization in the FabricPath network with a larger OSI Layer 2 domain. It also helps with simplifying service automation workflow by simply having less service dependencies to configure and manage.

Cloud Computing with the Titans

- Google
- Microsoft
- IBM
- AWS
- Salesforce

Accessing the Cloud

- Platform
  - Web Application frameworks
    - AJAX
    - Python Django
  - Web Hosting Service
    - Amazon Elastic Computing Cloud
    - Mosso
  - Proprietary Methods
    - Azure
    - Salesforce.com
- Web Applications
- Web APIs
- Web Browsers