

Hardware and Software for AI

Tushar B. Kute,
<http://tusharkute.com>



Hardware and Software

- Data Center
- Gateway Edge Computing
- Keyprocessor for AI
- CPU and GPU
- FiELd Programmable Gate Array (FPGA)

Data Center

- A data center is a building, a dedicated space within a building, or a group of buildings used to house computer systems and associated components, such as telecommunications and storage systems.
- Since IT operations are crucial for business continuity, it generally includes redundant or backup components and infrastructure for power supply, data communication connections, environmental controls (e.g., air conditioning, fire suppression), and various security devices.
- A large data center is an industrial-scale operation using as much electricity as a small town

Data Center



Data Center

- Modernization and data center transformation enhances performance and energy efficiency.
- Information security is also a concern, and for this reason, a data center has to offer a secure environment that minimizes the chances of a security breach.
- A data center must, therefore, keep high standards for assuring the integrity and functionality of its hosted computer environment.

Data Center

- Industry research company International Data Corporation (IDC) puts the average age of a data center at nine years old.
- Gartner, another research company, says data centers older than seven years are obsolete.
- The growth in data (163 zettabytes by 2025) is one factor driving the need for data centers to modernize.

Data Center

- Focus on modernization is not new: concern about obsolete equipment was decried in 2007, and in 2011 Uptime Institute was concerned about the age of the equipment therein.
- By 2018 concern had shifted once again, this time to the age of the staff: "data center staff are aging faster than the equipment."

Data Center: Types

- Data centers vary in size, from a small server room all the way up to groups of geographically distributed buildings, but they all share one thing in common: they are a critical business asset where companies often invest in and deploy the latest advancements in data center networking, compute and storage technologies.
- The modern data center has evolved from a facility containing an on-premises infrastructure to one that connects on-premises systems with cloud infrastructures where networks, applications and workloads are virtualized in multiple private and public clouds.

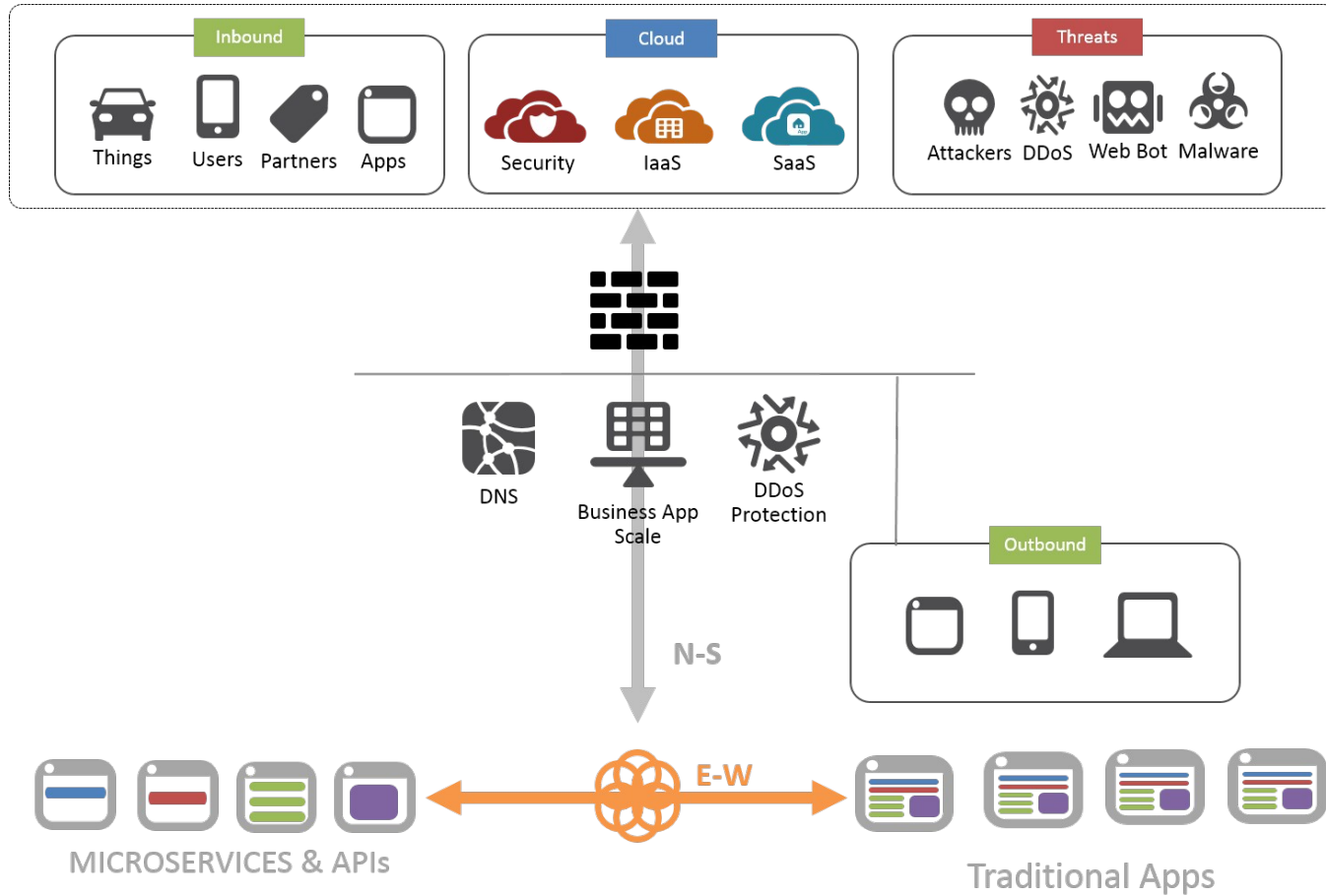
Data Center: Types

- Enterprise data centers are typically constructed and used by a single organization for their own internal purposes. These are common among tech giants.
- Colocation data centers function as a kind of rental property where the space and resources of a data center are made available to the people willing to rent it.
- Managed service data centers offer aspects such as data storage, computing, and other services as a third party, serving customers directly.
- Cloud data centers are distributed and are sometimes offered to customers with the help of a third-party managed service provider.

Evolution to Cloud

- The fact that virtual cloud DC can be provisioned or scaled-down with only a few clicks is a major reason for shifting to the cloud.
- In modern data centers, software-defined networking (SDN) manages the traffic flows via software.
- Infrastructure as a Service (IaaS) offerings, hosted on private and public clouds, spin up whole systems on-demand.
- When new apps are needed, Platform as a Service (PaaS) and container technologies are available in an instant.

Architecture



Architecture Components

- Data centers are made up of three primary types of components: compute, storage, and network.
- However, these components are only the top of the iceberg in a modern DC. Beneath the surface, support infrastructure is essential to meeting the service level agreements of an enterprise data center.

Data Center Computing

- Servers are the engines of the data center. On servers, the processing and memory used to run applications may be physical, virtualized, distributed across containers, or distributed among remote nodes in an edge computing model.
- Data centers must use processors that are best suited for the task, e.g. general purpose CPUs may not be the best choice to solve artificial intelligence (AI) and machine learning (ML) problems.

Data Center Storage

- Data centers host large quantities of sensitive information, both for their own purposes and the needs of their customers.
- Decreasing costs of storage media increases the amount of storage available for backing up the data either locally, remote, or both.
- Advancements in non-volatile storage media lowers data access times. In addition, as in any other thing that is software-defined, software-defined storage technologies increase staff efficiency for managing a storage system.

Data Center Networks

- Datacenter network equipment includes cabling, switches, routers, and firewalls that connect servers together and to the outside world.
- Properly configured and structured, they can manage high volumes of traffic without compromising performance.
- A typical three-tier network topology is made up of core switches at the edge connecting the data center to the Internet and a middle aggregate layer that connects the core layer to the access layer where the servers reside.
- Advancements, such as hyperscale network security and software-defined networking, bring cloud-level agility and scalability to on-premises networks.

Support Infrastructure

- Data centers are a critical asset that is protected with a robust and reliable support infrastructure made up of power subsystems, uninterruptible power supplies (UPS), backup generators, ventilation and cooling equipment, fire suppression systems and building security systems.
- Industry standards exist from organizations like the Telecommunications Industry Association (TIA) and the Uptime Institute to assist in the design, construction and maintenance of data center facilities.

Support Infrastructure

- For instance, Uptime Institute defines these four tiers:
 - Tier I: Basic capacity, must include a UPS.
 - Tier II: Redundant capacity and adds redundant power and cooling.
 - Tier III: Concurrently maintainable and ensures that any component can be taken out of service without affecting production.
 - Tier IV: Fault tolerant, allowing any production capacity to be insulated from ANY type of failure.

Security

- Data center firewalls, data access controls, IPS, WAF and their modern equivalent Web Application & API Protection (WAAP) systems need to be specified properly to ensure they scale as needed to meet the demands of data center networks.
- In addition, if you're choosing a data storage or cloud services provider, it's important that you understand the security measures they use for their own DC.
- Invest in the highest possible level of security to keep your information safe.

Thank you

This presentation is created using LibreOffice Impress 5.1.6.2, can be used freely as per GNU General Public License



@mitu_skillologies



/mITuSkillologies



@mitu_group



/company/mitu-
skillologies



MITUSkillologies

Web Resources

<http://mitu.co.in>

<http://tusharkute.com>

contact@mitu.co.in

tushar@tusharkute.com