MODULE 2 · STUDY NOTES

Data Center Physical Infrastructure - Overview

The seven elements of network-critical physical infrastructure, the four operating environments, and the best practices and challenges that keep a data center running 24/7

A data center is often pictured as racks of servers, switches, and storage — but that visible IT infrastructure only works because of the layer beneath it. That layer is the **Network-Critical Physical Infrastructure (NCPI)**: the power, cooling, racks and physical structure, cabling, security and fire protection, management, and services that together form the base of the IT infrastructure pyramid. Because failure of one element can bring down the whole data center, NCPI is treated as the most important aspect of running a data center successfully. This module maps the seven elements in detail, reframes them as the four key operating environments a data center must provide — power, cooling, IT connectivity, and space — and then turns practical: the best practices that maximize availability and reliability and increase MTBF, and the modern challenges (from data security to IoT) that operators must constantly manage. The recurring theme is redundancy and disciplined operations — the right infrastructure, designed for growth and maintained by skilled staff, is what delivers the 24/7 reliable service the business depends on.

LEARNING OBJECTIVES

- Identify the key elements of a Network-Critical Physical Infrastructure (NCPI) for a data center
- Describe the key elements of an NCPI for a data center
- Recognize the importance of an NCPI in a data center
- Recognize the role of the four key environments in a data center
- Recognize the importance of best practices associated with the seven NCPI elements
- Identify the best practices associated with the seven NCPI elements
- Identify common challenges experienced during the implementation of an NCPI
- Recognize how issues associated with NCPI elements can impact a data center

2.1 The seven elements of NCPI

NCPI is the layer below the physical infrastructure — the servers and switches — in the data center, and it is seen as the most important aspect of running a data center successfully. The NCPI layer is made up of elements that are each essential in their own right. The **seven elements** are power, cooling, racks and physical structure, cabling, security and fire protection, management, and services. All seven together form the base of the IT infrastructure pyramid, which means that failure of one or more NCPI elements can lead to the potential failure of the data center. For this reason, a reliable NCPI system must be set up to run a reliable business.

TABLE 2.1 The seven NCPI elements and their function

ELEMENT	FUNCTION
Power	Current from the grid or battery, distributed through transformers and distribution panels
Cooling	Precision air conditioners, cooling towers, and condensers used to cool the equipment
Racks and physical structure	IT racks to house the equipment, plus dropped ceilings, raised floors, and pathways that manage cabling
Cabling	Data and power cables that form the cabling infrastructure of the data center
Security and fire protection	Remote monitoring and biometric identification, plus fire detection and suppression
Management	Overseeing enterprise-level systems that monitor buildings and maintain data center health
Services	Meeting the user's expectation of data center service with 24/7 availability

Power. A reliable power infrastructure includes a range of elements that support the electrical services from the building entrance to the data center loads. Power is supplied to a data center through a connection to the main utility grid. Should the grid experience an outage, the UPS provides a temporary source of power, and if the outage is sustained the building's backup generators engage to provide a steady source until utility power recovers. The **low voltage (LV) switchgear** distributes low-voltage electricity to endpoints such as the UPS, and transformers step down high-voltage grid power to a voltage the equipment can use.

EQUIPMENT IN THE POWER ELEMENT

- **Generators** — at least one backup generator per data center
- **Transformers** — step high-voltage grid power down to usable voltage
- **Power distribution unit (PDU)** and **power distribution panels**
- **Uninterruptible power supply (UPS) systems and batteries**
- **Surge protection** and **circuit breakers**

WORKED EXAMPLE – RIDING THROUGH A POWER OUTAGE

Utility power is flowing normally to the data center. The grid then fails. The UPS immediately supplies short-term power from its batteries. The backup generator usually takes **about 1 minute** to start up, and during that minute the UPS carries the load. After the generator starts, the load is transferred to the generator, generator power is supplied through the UPS, and the batteries begin to recharge. The generator continues to supply power until the utility power is confirmed as stable. This hand-off — UPS first, generator second — is why a data center can survive both a momentary flicker and a sustained outage.

Cooling. Cooling systems are essential for managing temperature and humidity within the data center so the IT equipment functions properly. Data centers use a lot of power, and that power generates a lot of heat; the cooling system's components maintain a constant environment and prevent hot zones and heat spillovers. Two processes cool the equipment: **cooling the hot air generated within the data center**, and **pumping outside air into the data center during ambient climatic conditions.**

COOLING SYSTEM COMPONENTS

- **Chillers** (air or water cooled) and **cooling towers**
- **Computer room air handlers (CRAHs)**
- **Computer room air conditioners (CRACs)** and external condensers
- **Ductwork**, **pump sets**, and flow and return piping
- **Air grills** (duct and floor)

Racks and physical structure. Racks are a critical component of a data center's physical infrastructure, constructed using steel frames to hold the electrical equipment. Intelligent data center racks are now widely used to improve performance and combat critical failures. Proper maintenance of organized racks is vital: it lets data center administrators quickly resolve critical issues by identifying the source of failure.

COMMON PITFALL

Do not leave excessive open space in a rack. Equipment in the racks generates heat, and open spaces allow hot air to recirculate inside the rack, which increases the heat and reduces the performance of the equipment. If racks are not maintained properly, the result can be acute equipment failure and heavy financial loss. Instead, organize equipment without excessive gaps and monitor rack loads for the heat they generate, so personnel can determine and eliminate hot pockets by distributing the rack loads.

Cabling. The cabling infrastructure is fundamental to supporting the data center because it holds the entire data networking and power cables; good cabling designs also hold the cable trays and cable management devices. Two problems come up repeatedly: **wrong plug-in of cables due to human error**, and **overheating of cables due to a sudden surge in power supply.** A proper cabling infrastructure lessens the probability of downtime caused by human error and high temperature.

2.2 Security, management, and services

Security and fire protection systems are vital in upholding the reliability, security, and readiness of the data center. Their subsystems span both the facility level and the rack level, and they combine two jobs: protecting data and physical assets, and protecting against fire through detection and suppression.

SECURITY AND FIRE PROTECTION SYSTEMS

- Fire-resistant floors and roofs; automatic fire detection units
- Automatic fire protection systems — sprinklers or gaseous clean agent extinguishing systems
- Portable extinguishers and hose lines; water mist protection systems
- Intelligent smoke detectors and linear heat detection
- Firewalls and data security applications
- Biometric devices, security cameras, and codes and access cards

Hazards are real: water leakage, accidental spills, inert gases, or any type of chemicals can create a hazardous environment, increasing the risk of endangering people and equipment and even causing fire and explosions. Although fire safety is important, it is equally vital to protect the facility from intruders, because data or equipment theft can be just as detrimental. Data centers control physical access using biometric devices, security cards, cameras, and access cards to govern who enters.

TABLE 2.2 Three compliance levels for fire detection and communication

LEVEL	REQUIREMENT
System level	Fire detection and communication practice must comply with industry standards
Component level	Compliance must ensure the use of correct suppression agents approved by industry practices
Installation level	Correct procedures for installation must be followed

Management. NCPI management oversees the issues and challenges that occur within the NCPI environment, and it is vital for ensuring the NCPI elements are reliable at all times. Management includes **building management systems (BMS)**, **network management systems (NMS)**, **element managers** (such as APC InfraStruXure manager), and other monitoring hardware and software.

Services. Data center management must perform certain services to sustain the NCPI systems while they are operational. These include consulting and design, managed outsourcing, user authentication and authorization management, load balancing management, controlled access to the internet, power distribution and backup power management, hardware installation, technical support, data backup and archiving, decommissioning, maintenance and repair, and monitoring. These services are performed by skilled personnel — if the required skill set is not available or adequate training is not provided, human errors can occur during day-to-day operations and cause problems.

Why NCPI matters. All seven elements are important and should integrate effortlessly to meet the increasing demand of data centers, and their management plays a role in the overall functioning of the data center. To sustain an exceptionally reliable and available data center, operators maintain **redundant systems** for the important elements of the framework. NCPI does not stand alone: alongside it, the **IT services** that keep business operations running — storage, data processing servers, and communication systems — store and process the data. To manage both the NCPI elements and the IT equipment, **standardized processes** should be followed by all data center employees; failing to implement them can lead to operational failure.

2.3 The four key environments

A data center — small or large, serving a single organization or many — must provide **four key environments** for successful operation: **power, cooling, IT connectivity, and space.** These environments need sufficient and redundant power arrangement, adequate cooling, proper IT connectivity, and enough space for the quantity of servers and switches. The four environments are the operational lens on the same infrastructure the seven elements describe.

TABLE 2.3 The four key environments and their role

ENVIRONMENT	ROLE IN THE DATA CENTER
Power	Provides continuous availability in the event of disasters or even a major grid failure
Cooling	Manages temperature and humidity to help the IT equipment function properly
IT connectivity	Maintains network service within the data center and also with external customers
Space	Ensures effective storage types, server efficacy, ideal rack layout, and cabling design

Power. A reliable power supply is the most important aspect of continuous availability. Power is sourced from the main grid of a local utility company — many data centers connect **two separate transformers** to the grid to maintain redundancy. **Step-down transformers** reduce the high-voltage grid supply to a lower voltage. There should be at least one **diesel or gas backup generator** to avoid a large outage; these generators have the capacity to run for about **48**

hours and are sized to the equipment they must support. For continuous supply, **UPS systems** provide temporary power until the generators start. The power system also includes **main distribution boards (MDBs)**, which house fuses, circuit breakers, and ground leakage protection units and distribute incoming low-voltage electricity to endpoints across the data center.

Cooling. As more data is stored, processed, and distributed, computing power increases, which raises the power density in data centers and drives an increased need for efficient cooling. Cooling is a **key performance indicator (KPI)** for data center operations — the absence of a proper cooling mechanism impacts reliability. Cooling is achieved by adopting various technologies, each suited to different conditions.

TABLE 2.4 Cooling technologies and their applications

TECHNOLOGY	APPLICATION
Calibrated vectored cooling (CVC)	Used for high-density servers; manages heat by optimizing the airflow between equipment
Chilled water system	Used in mid- to large-size data centers; uses chilled water to cool the air
Cold aisle / hot aisle design	Racks arranged in alternating aisles — cold aisles take in cold air at the front, hot aisles exhaust hot air from the back
Free cooling	Uses outside atmospheric air rather than continually chilling the same air; suited to specific climates and very energy efficient
Raised floor	A platform on the concrete slab that gives space for water-cooling pipes or increased airflow

IT connectivity covers two aspects: **connectivity**, which describes how well hardware or software devices can communicate with a range of other devices, and **networks**, which provide a common map for servers and other equipment to run the workload and exchange data. Setting up connectivity means understanding the network components, the protocols that move data between devices, and newer technologies.

NETWORK COMPONENTS

- **Servers**
- **Switches** that connect devices together
- **Routers** for packet forwarding
- **Controllers** to manage the workflow between network devices
- **Gateways** for connecting the data center network to the internet
- As a proactive redundancy measure, use multiple redundant fiber connections to multiple internet service providers (ISPs) so the data center is always connected

TABLE 2.5 Data center networking protocols

PROTOCOL	ROLE
TCP/IP	Transport Control Protocol/Internet Protocol — allows routers to pass packets of data to each other
Ethernet	Connects devices using media access control (MAC) addresses
VXLAN	Virtual Extensible LAN — supports an overlay network on top of the physical network, enabling virtualization of network equipment and making cloud computing possible
OpenFlow	The underlying protocol of software-defined networking (SDN); helps a network controller determine the path of data flow

Newer technologies. Software-defined networking (SDN) efficiently accommodates the varying workloads on servers. Its architecture separates a router's **control plane** from the **data (forwarding) plane**: the control plane makes routing decisions and instructs the network to accommodate data, while the data plane manages media and content and forwards packets through the router. These networking technologies are essential for easy execution and communication between devices when managing large volumes of data.

Space. Selecting a site requires proper analysis of location, availability of power, and proper access, plus awareness of local data protection laws, tax rules, and availability of labor. Designing the data center must anticipate **additional space for future expansion** — adding more racks and cabinets — and account for the use of space, the type of racks, cooling systems, distribution of cables, and power points. Standards such as **TIA-942** should be followed so that servers and storage can be added or upgraded with minimal disruption, both today and tomorrow.

2.4 Best practices for the seven NCPI elements

The goal of a data center is to provide 24/7 reliable services to the companies that use it, and to do so it should be **flexible, scalable, and offer long-term sustainability.** Recent advances in computing and networking have made operations more complex, so following standard processes with skilled personnel — and incorporating best practices for the NCPI elements — becomes extremely important. Without them, a data center risks operational failure.

WHAT THE ELEMENTS SHOULD BE DESIGNED AND MAINTAINED FOR

- **Designed for:** maximum availability, maximum reliability, reduced operation downtime, concurrent maintainability, and low costs (including energy and operating costs)
- **Regularly maintained for:** increasing the mean time between failures (MTBF)
- **Benefits of best practices:** improving availability, improving reliability, optimizing MTBF, improving mean time to repair (MTTR), and containing the total cost of operations

KEY TERMS

Availability
The readiness of systems to be accessible when needed — the probability that a system is running and operational at a specific point of time or as a percentage of a time interval

Reliability
The probability that a system will perform its required function at a specific point of time or over a period of time

MTBF (mean time between failures)
The average expected time between two failures of a system

MTTR (mean time to repair)
The average time required to repair a system

Improving availability. The availability of a system can be increased by incorporating a **standby redundant system or resource.** Each NCPI element has its own best practices, but the principle is the same: keep a backup ready so a failure does not take the service down.

TABLE 2.6 Best practices to improve availability, by element

ELEMENT	BEST PRACTICE FOR AVAILABILITY
Power	Add a backup power system such as a fuel-run generator, plus backup fuel beyond the fuel stored in the generator
Cooling	Provide 2N redundant cooling capacity, link backup power to the cooling systems, and use equipment that starts quickly after a power outage
Racks	Plan for future capacity, use rack space properly so cooling and airflow are maintained, and monitor rack power to avoid energy waste
Cables	Follow the guidelines of ANSI, TIA, ANSI/TIA-942, and IEEE
Fire suppression	Perform regular maintenance and check that the devices are working normally
Management and services	Keep a consistent employee base, train regularly, follow standard procedures (documenting deviations), and provide backup procedures with a failover system

A note on the cooling entry: **2N cooling capacity** means the cooling system has two sets of chillers, two sets of pumps, two sets of pipework, and two CRAHs. If one set of any component fails, the other set takes over — which is exactly what keeps cooling available during a fault.

Improving reliability. Improving reliability is not limited to the design of the products and equipment themselves. Reliability is improved by **performing regular maintenance** and **closely monitoring major components and replacing their parts before they wear out.** It is further improved by keeping the operational load within the product-specified limit, and, when a failure occurs, by assessing the probable cause — mechanical stress, fatigue, friction, defective parts, or maintenance-induced error — and the impact and severity of each failure for the entire system. For power, cooling, cabling, and fire suppression, the data center should have **redundant components** so the probability of the entire system failing is low. For management and services, reliability is maintained by allocating operations to keep the system effective.

COMMON PITFALL

Availability and reliability are not the same, and they are improved differently. **Availability** answers "is the system operational right now?" and is raised with standby redundancy — backup fuel, 2N cooling, fast-start equipment. **Reliability** answers "will the system keep performing its required function over time?" and is raised with better design, regular maintenance, and replacing worn parts early. Improving one does not automatically improve the other.

2.5 Challenges and their impact

The computing, network, and storage resources of a data center rely on its physical infrastructure to function as intended — which is why those physical components are termed NCPI elements. An **inadequate physical infrastructure can create serious problems** later: cabinets and racks must be planned for future expansion, and security and fire protection must be designed to protect the network at the building, room, and rack levels. The right physical infrastructure must therefore be **flexible in adapting to future business requirements**, realigning NCPI components — updating legacy systems or adding equipment — whenever business needs change.

MODERN ISSUES AFFECTING THE NCPI ELEMENTS

- **Data storage** — adding equipment to existing racks must factor in the overall heat generated
- **Cooling** — the cooling system must be fine-tuned or upgraded to handle the added heat
- **Power and load balancing** — power distribution and load balancing for new equipment must be reevaluated and corrected
- **Virtualization** — servers are mapped to multiple virtual cloud servers; fewer physical servers are used, but this increases power consumption
- **Real-time monitoring** — operators must know everything happening in the data center as it happens (for example, a heat pocket must trigger an alert or an automatic cooling response)
- **Management tools** — monitoring and alert systems require investment in intelligent monitoring tools

TABLE 2.7 Six common NCPI implementation challenges

CHALLENGE	WHAT IT INVOLVES	ELEMENTS AFFECTED
Data security	Secure critical equipment from physical and cyber theft	Security & fire protection
Cost control vs. efficiency	Manage the cost of high availability, reliability, resilience, and real-time monitoring tools	Power and cooling
Capacity planning	Ensure deployed resources (IT equipment, power, cooling, support) meet the workload demand	All seven
Power management	Manage increased power demand as storage and operation technologies evolve	Power and cooling
IoT	Manage the huge volume of data generated by monitoring devices in the data center	All seven
Mobile enterprise	Build the capability to present data across all platforms and devices	All seven

The same six categories are examined as **impacts** on the data center. **Data security** is one of the largest ongoing challenges: a data breach can cost millions of dollars and poses risks of intellectual property loss, exposure of confidential data, loss of personally identifiable information, and misuse of data. **Balancing cost control and efficiency** adds cost pressure, because the equipment requires real-time monitoring, security, fire suppression, and backup power, and some of it must be made redundant — costs offset through environmental regulations (such as renewable solar energy to reduce carbon footprint) and virtualization (sharing physical servers to emit less heat and lower operating cost).

Capacity planning means running at peak capacity while leaving a **capacity safety gap** for the unexpected; overprovisioning that gap wastes storage, processing, cooling, and power, so many data centers use a **DCIM** system to identify unused cooling capacity and storage and run at full capacity while minimizing risk. **IoT** — an estimated 50 billion connected devices — produces a tremendous amount of data that data centers must process, prioritize, store, and analyze, forcing them to evolve. **Mobile enterprise** devices offer immediate access to business-critical data, so they must be monitored, secured, and restricted, with new safeguards able to remotely format or track a lost or stolen device while safeguarding user data and privacy.

> **WORKED EXAMPLE – WHY POWER MANAGEMENT NEVER ENDS**
>
> To ensure high availability, administrators deploy redundant servers — but each redundant server still needs power, backup power, security, maintenance, and monitoring, and it generates heat, so cooling for that part of the data center must rise too. Strategies meant to help can backfire: server consolidation and virtualization optimize hardware but **increase** power consumption, and modern blade servers can consume **four to five times** the energy of previous-generation storage devices. On top of that, **cooling accounts for about 40 percent of a data center's average power consumption.** Adding capacity therefore compounds both the power and the cooling demand, which is why power and cooling management is an ongoing challenge for data center administrators.

Module summary

- NCPI is the layer below the physical IT infrastructure that is essential to running a data center; its seven elements — power, cooling, racks and physical structure, cabling, security and fire protection, management, and services — form the base of the IT infrastructure pyramid, so failure of one can fail the data center.
- The power element combines the utility grid, step-down transformers, generators, UPS and batteries, PDUs and distribution panels, surge protection, and circuit breakers; on an outage the UPS bridges power for about a minute until the generator starts and takes the load.
- Cooling manages temperature and humidity to prevent hot zones and heat spillovers using chillers, cooling towers, CRACs/CRAHs, condensers, ductwork, pumps, and piping; racks must be organized without open gaps, and cabling must avoid wrong plug-ins and overheating.
- Security and fire protection span cyber and physical protection (firewalls, biometrics, cameras, access cards) and fire detection and suppression, with compliance required at the system, component, and installation levels; management uses BMS, NMS, and element managers, and services keep the NCPI running with skilled staff.
- The four key environments — power, cooling, IT connectivity, and space — reframe the same infrastructure operationally; power needs redundant transformers, ~48-hour generators, UPS, and MDBs, and cooling uses CVC, chilled water, hot/cold aisle, free cooling, and raised floors.
- IT connectivity relies on servers, switches, routers, controllers, and gateways, redundant fiber to multiple ISPs, TCP/IP and Ethernet plus VXLAN and OpenFlow/SDN; space selection and design follow TIA-942 and plan for future expansion.
- Best practices design elements for maximum availability, reliability, reduced downtime, concurrent maintainability, and low cost, and maintain them to increase MTBF (mean time between failures) and improve

MTTR (mean time to repair); availability — the probability a system is operational at a point in time — is raised with standby redundancy (backup fuel, 2N cooling, fast-start equipment, ANSI/TIA/IEEE cabling standards), while reliability — the probability it performs its function over time — is raised with better design, maintenance, and early parts replacement.

- Six challenges dominate NCPI implementation — data security, cost control vs. efficiency, capacity planning, power management, IoT, and mobile enterprise — several touching all seven elements; DCIM helps run at full capacity, and power and cooling demand keeps rising because cooling alone is about 40 percent of average power consumption.

Review questions

1. **What is NCPI, and what are its seven elements?**

   Answer: NCPI is the layer below the physical IT infrastructure (servers, switches, storage) that is essential to running a data center. Its seven elements are power, cooling, racks and physical structure, cabling, security and fire protection, management, and services.

2. **Trace what happens to data center power from a normal utility feed through a sustained outage.**

   Answer: Normally the utility grid supplies power through step-down transformers. On an outage the UPS supplies short-term battery power; the backup generator takes about 1 minute to start; once running, the load transfers to the generator (supplied through the UPS) and the batteries recharge; the generator runs until utility power is confirmed stable.

3. **Why must rack equipment be organized without excessive open spaces?**

   Answer: Open spaces let hot air recirculate inside the rack, which increases the heat and reduces equipment performance. Monitoring rack loads for the heat they generate helps identify and eliminate hot pockets by distributing the loads.

4. **Name the four key environments and one distinguishing fact about each.**

   Answer: Power — continuous availability through disasters or a grid failure. Cooling — temperature and humidity control. IT connectivity — network service internally and to external customers. Space — effective storage, rack layout, and cabling design, following TIA-942.

5. **Distinguish availability, reliability, MTBF, and MTTR.**

   Answer: Availability is the probability a system is operational at a point in time; reliability is the probability it performs its required function at a time or over a period; MTBF is the average expected time between two failures; MTTR is the average time required to repair a system.

6. **How is availability improved for power, cooling, and cables specifically?**

   Answer: Power — a backup generator plus extra backup fuel. Cooling — 2N redundant capacity, backup power linked to cooling, and fast-start equipment. Cables — follow the ANSI, TIA-942, and IEEE standards.

7. **Name the six common NCPI implementation challenges.**

   Answer: Data security, balance between cost control and efficiency, capacity planning, power management, IoT, and mobile enterprise.

8. **Why is power management described as an ongoing challenge?**

   Answer: Equipment keeps evolving and drawing more power, and server consolidation and virtualization can increase rather than reduce it. Blade servers can use four to five times the energy of previous-generation storage, and cooling is about 40 percent of average power consumption — so added capacity compounds both power and cooling demand.