

316

Storage Technology

- Course contents

- Introduction to Information Storage.
- Data Center Environment.
- Data Protection: RAID
- Intelligent Storage Systems.
- Fiber Channel Storage Area Networks.
- IP SAN and FCoE.
- Network-Attached Storage.

- **Foreword**

- Data carries information during the transmission on networks. What is the relationship between information and data? What is the function of data storage? This course describes the definition of information and data in the computer field, their relationship, as well as the concept, development history, and development trend of data storage.

Objective

- On completion of this course, you will be able to understand:
 - Definition of information and data
 - Concept of data storage
 - History of data storage
 - Development trend of data storage products



Contents

1. Data and Information

2. Data Storage

3. Development of Storage Technologies

4. Development Trend of Storage Products

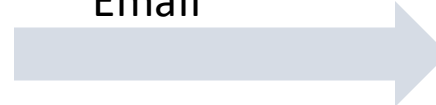
What is Data

SNIA (Storage Networking Industry Association) defines data as the digital representation of anything in any form.

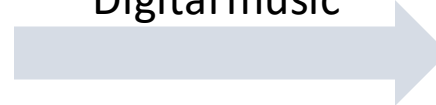
- Data as a general concept refers to digits, letters, and symbols that can be input into a computer and processed by a computer program.
- DAMA defines data as the expression of facts in the form of texts, numbers, graphics, images, sounds, and videos.
- SNIA is short for Storage Networking Industry Association.
- DAMA refers to the Global Data Management Community.



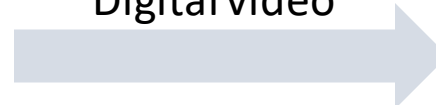
Email



Digital music



Digital video



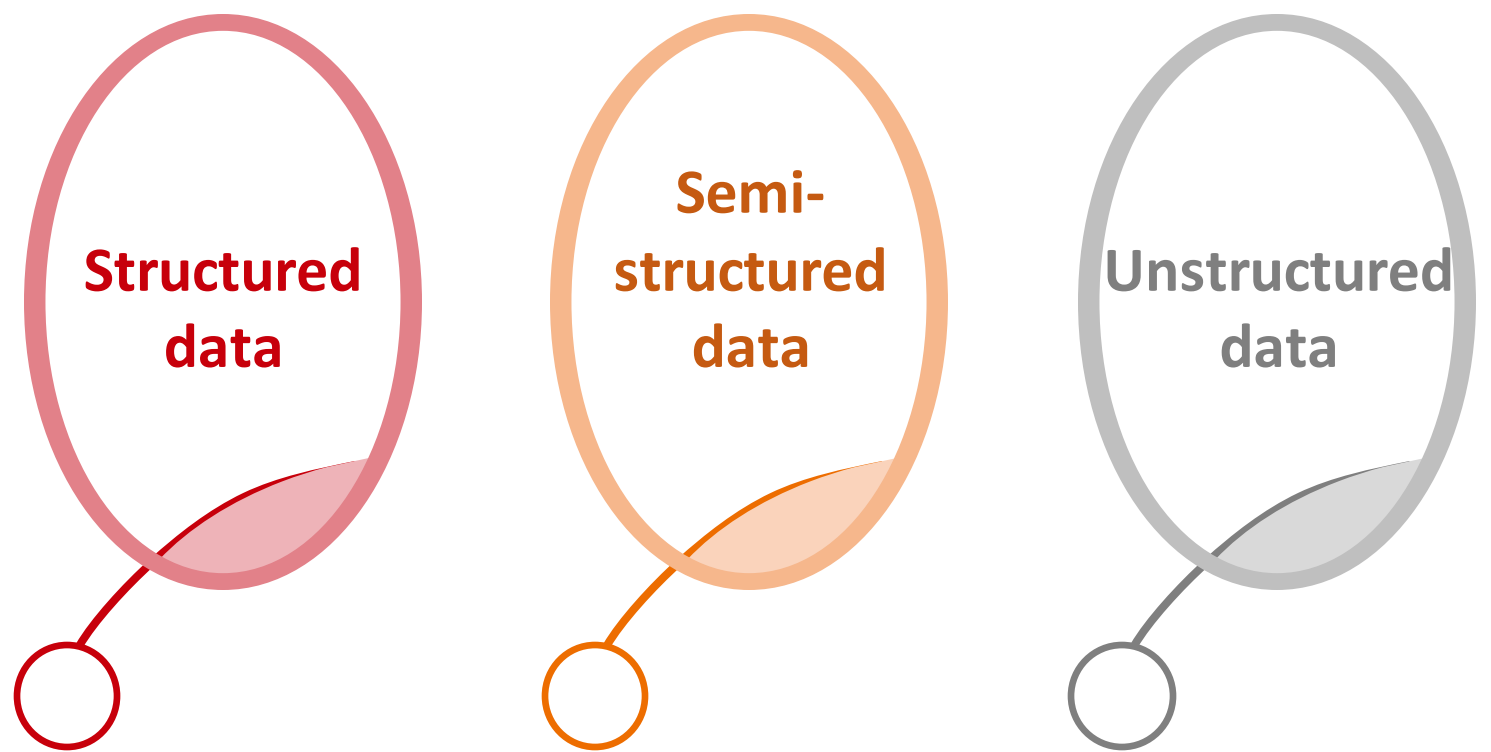
Ebook



Format in which data is stored

```
0101001010100010000011
1100011100010001110001
1100000111101010100101
0101001010100101001010
1001010101001010100010
1010010101001010101010
0101010101010010100010
0101001010101010100101
0101001010101010100101
0101010010100101001000
1010101001001010010010
```

Data Types



Based on data storage and management modes, data is classified into structured, semi-structured, and unstructured data.

Structured data

It can be represented and stored in a relational database, and is often represented as a two-dimensional table.

SQL server, MySQL, Oracle

Semi-structured data

It does not conform to the structure of relational databases or other data tables, but uses tags to separate semantic elements or enforces hierarchies of records and fields.

XML, HTML, JSON

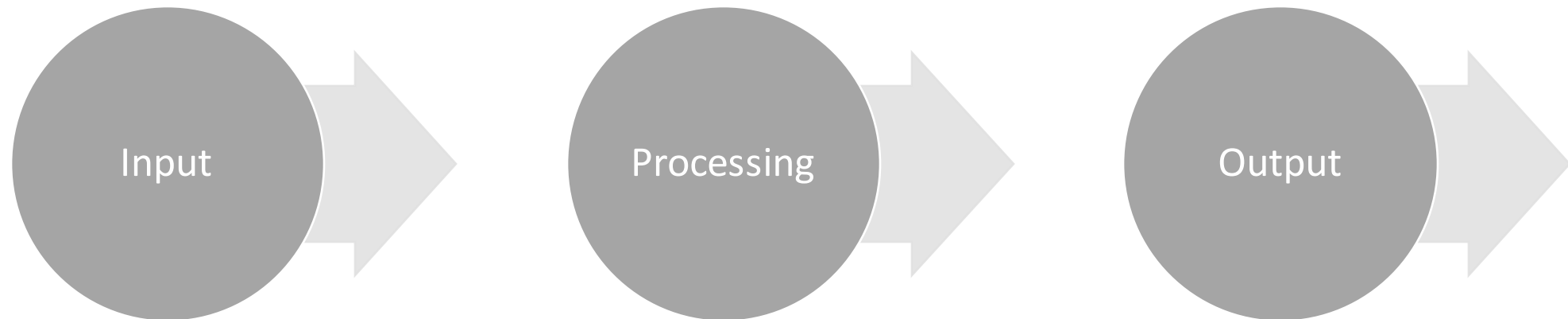
Unstructured data

It is not organized in a regular or complete data structure, or does not have a predefined data model.

Texts, pictures, reports, images, audios, and videos

Data processing Cycle

Data processing is the reorganization or reordering of data by humans or machines to increase their specific value. A data processing cycle includes three basic steps: input, processing, and output.



- Input: inputs data in a specific format, which depends on the processing mechanism. For example, when a computer is used, the input data can be recorded on several types of media, such as disks and tapes.
- Processing: performs actions on the input data to obtain more data value. For example, the time card hours are calculated to payroll, or sales orders are calculated to generate sales reports.
- Output: generates and outputs the processing result. The form of the output data depends on the data use. For example, the output data can be an employee's salary.

What Is Information

Information is processed, structured, or rendered in a given context to make it meaningful and useful.

Information is processed data, including data with context, relevance, and purpose. It also involves the manipulation of raw data.

- There are many definitions of information. Most of them in the computer field are generated based on the definitions proposed by Claude Elwood Shannon, known as "the father of information theory." His theory depicts that the essence of information is the resolution of random uncertainty. This means that the information is useful.

Data vs. Information

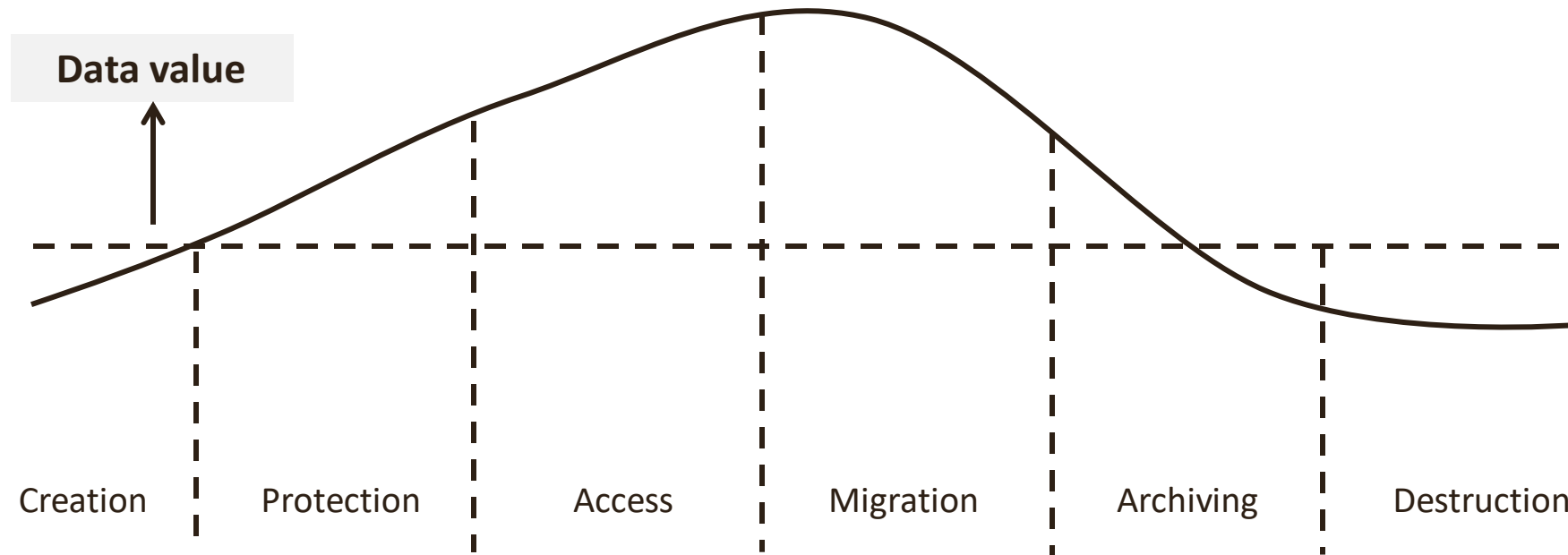
After being processed, data can be converted into information. In order to be stored and transmitted in IT systems, information needs to be processed as data.

Item	Data	Information
Feature	Raw and meaningless, with no specific purpose	Valuable and logical
Essence	Original materials	Processed data
Dependence	Data never depends on information	Information depends on data
Example	Meteorological data or satellite image data	Weather forecasts

- Data is a raw, unorganized data bit that needs to be processed to make it meaningful, whereas information is a set of data that is processed, interpreted, and presented to become meaningful, in accordance with the given requirement.
- Data does not have any specific purpose whereas information carries a meaning that has been assigned by interpreting data.
- Data alone has no significance while information is significant by itself.
- Data never depends on information while information is dependent on data.
- Data is measured in bits and bytes while information is measured in meaningful units like time and quantity.
- Data can be structured, tabular data, graph, data tree whereas information is language, ideas, and thoughts based on the given data.

Information Lifecycle Management

Information lifecycle management (ILM) refers to a set of management theories and methods from the stage when the information is generated and initially stored to the stage when the information is deleted.



Data creation phase: Data is generated from terminals and saved to storage devices.

Data protection phase: Different data protection technologies are used based on data and application system levels to ensure that various types of data and information are effectively protected in a timely manner. A storage system provides data protection functions, such as RAID, HA, disaster recovery (DR), and permission management.

Data access phase: Information must be easy to access and can be shared among organizations and applications of enterprises to maximize business value.

Data migration phase: When using IT devices, you need to upgrade and replace devices, and migrate the data from the old to new devices.

Data archiving phase: The data archiving system supports the business operation for enterprises by providing the record query for transactions and decision-making. Deduplication and compression are often used in this phase.

Data destruction phase: After a period of inactivity, data is no longer saved. In this phase, it is normal to destroy or reclaim data that does not need to be retained or stored, and clear the data from storage systems and data warehouses.

Contents

1. Data and Information

2. Data Storage

3. Development of Storage Technologies

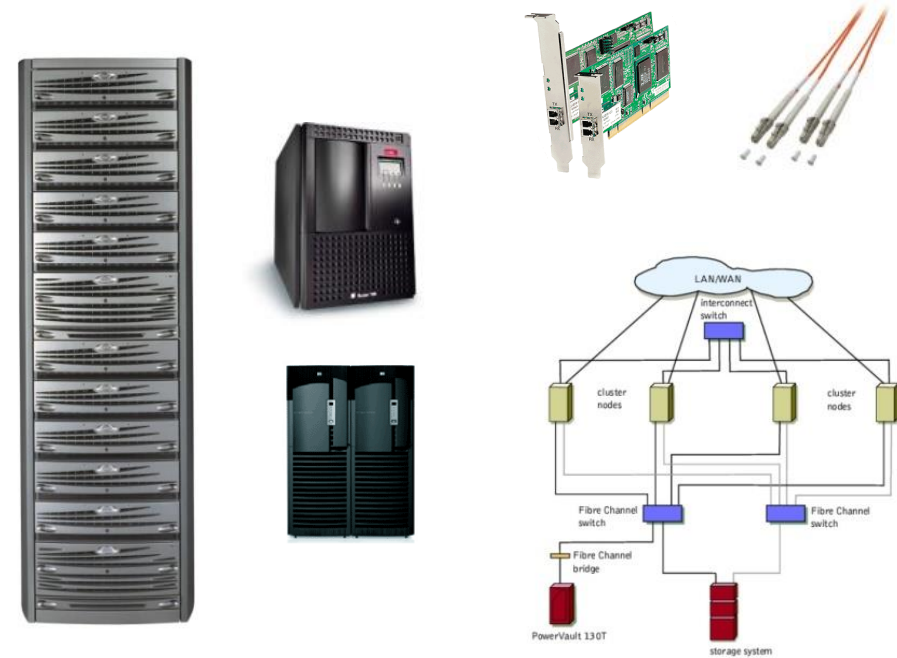
4. Development Trend of Storage Products

What is data storage?

Storage in a narrow sense



Storage in a broad sense

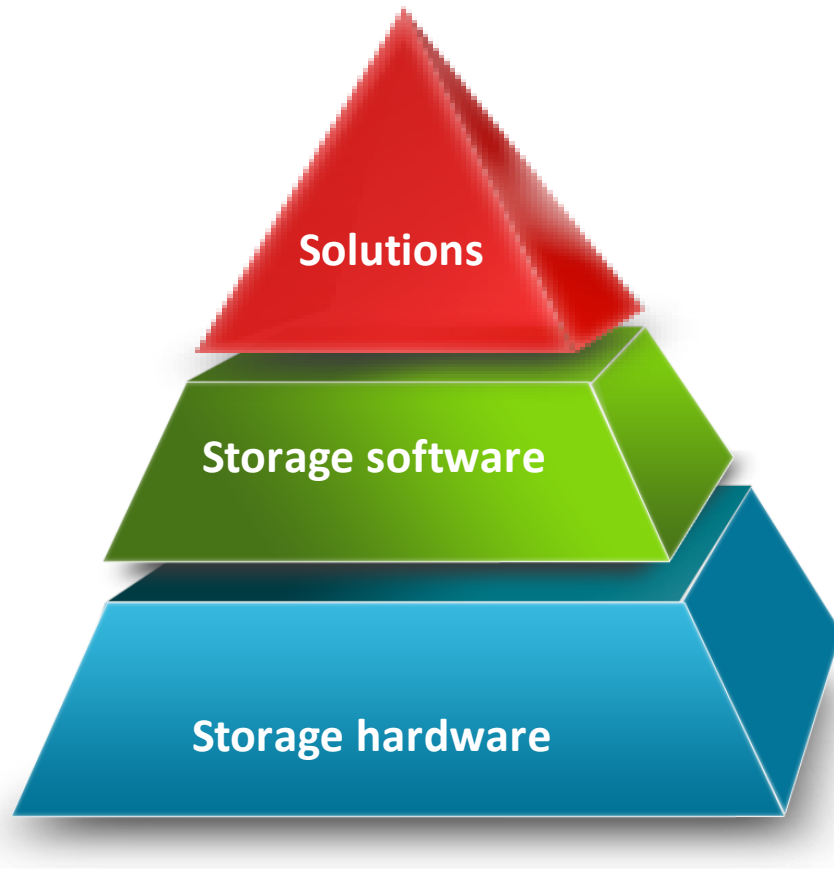


In a narrow sense, storage refers to the physical storage media with redundancy, protection, and migration capabilities, such as floppy disks, CDs, DVDs, disks, and even tapes.

In a broad sense, storage refers to a portfolio of solutions that provide enterprises with information access, protection, optimization, and utilization. It is the pillar of the data-centric information architecture.

Data storage covered in this course refers to storage in a broad sense.

Data Storage System



- Disaster recovery (DR) solutions
 - Backup solutions
-

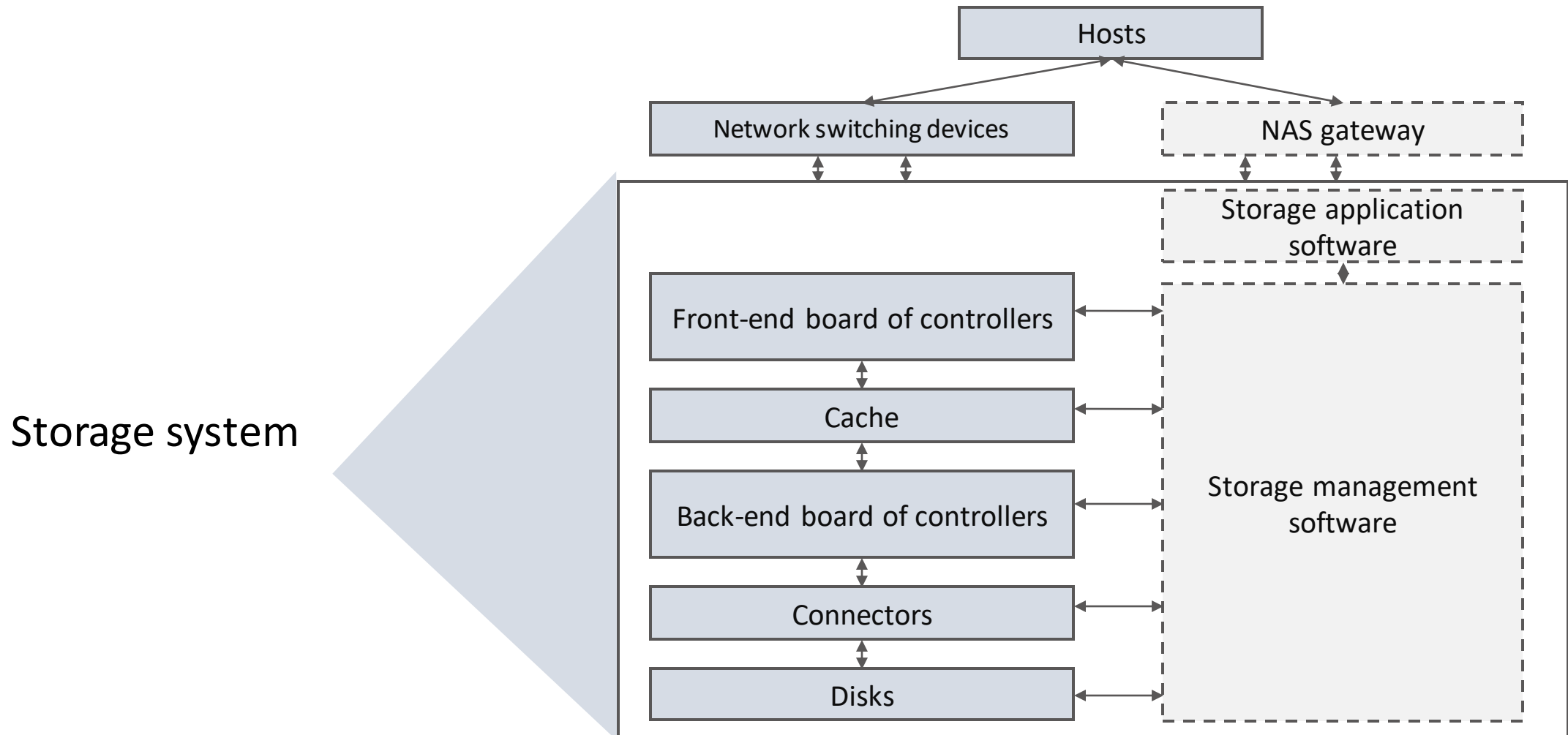
- Storage management software
 - Snapshot and mirroring software
 - Backup software
 - Multipathing software
-

- **Storage devices**
 - Disk array
 - Tape library
 - Virtual tape library
 - ...

Storage technologies are not separate or isolated. Actually, a complete storage system consists of a series of components.

A storage system consists of storage hardware, storage software, and storage solutions. Hardware involves external storage systems and devices for storage connections, such as disk arrays, tape libraries, and Fiber Channel switches. Storage software greatly improves the availability of a storage device. Data mirroring, data replication, and automatic data backup can be implemented by using storage software.

Physical structure of storage



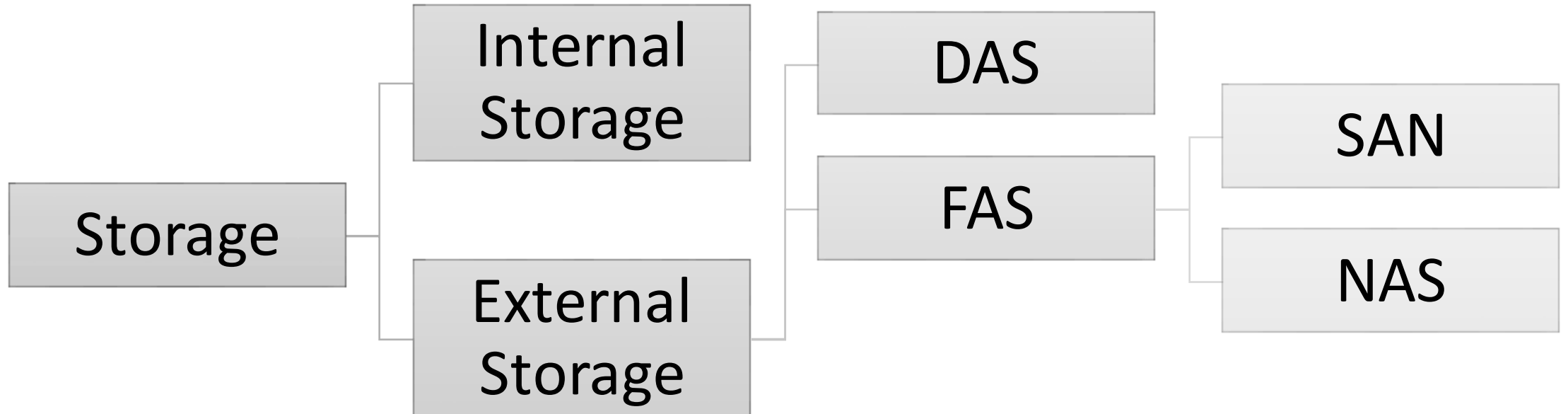
Physical structure of Storage

A typical storage system comprises the disk, control, connection, and storage management software subsystems. In terms of its physical structure, disks reside in the bottom layer and are connected to back-end cards and controllers of the storage system via connectors such as optical fibers and serial cables.

The storage system is connected to hosts via front-end cards and storage network switching devices to provide data access services.

Storage management software is used to configure, monitor, and optimize subsystems and connectors of the storage system.

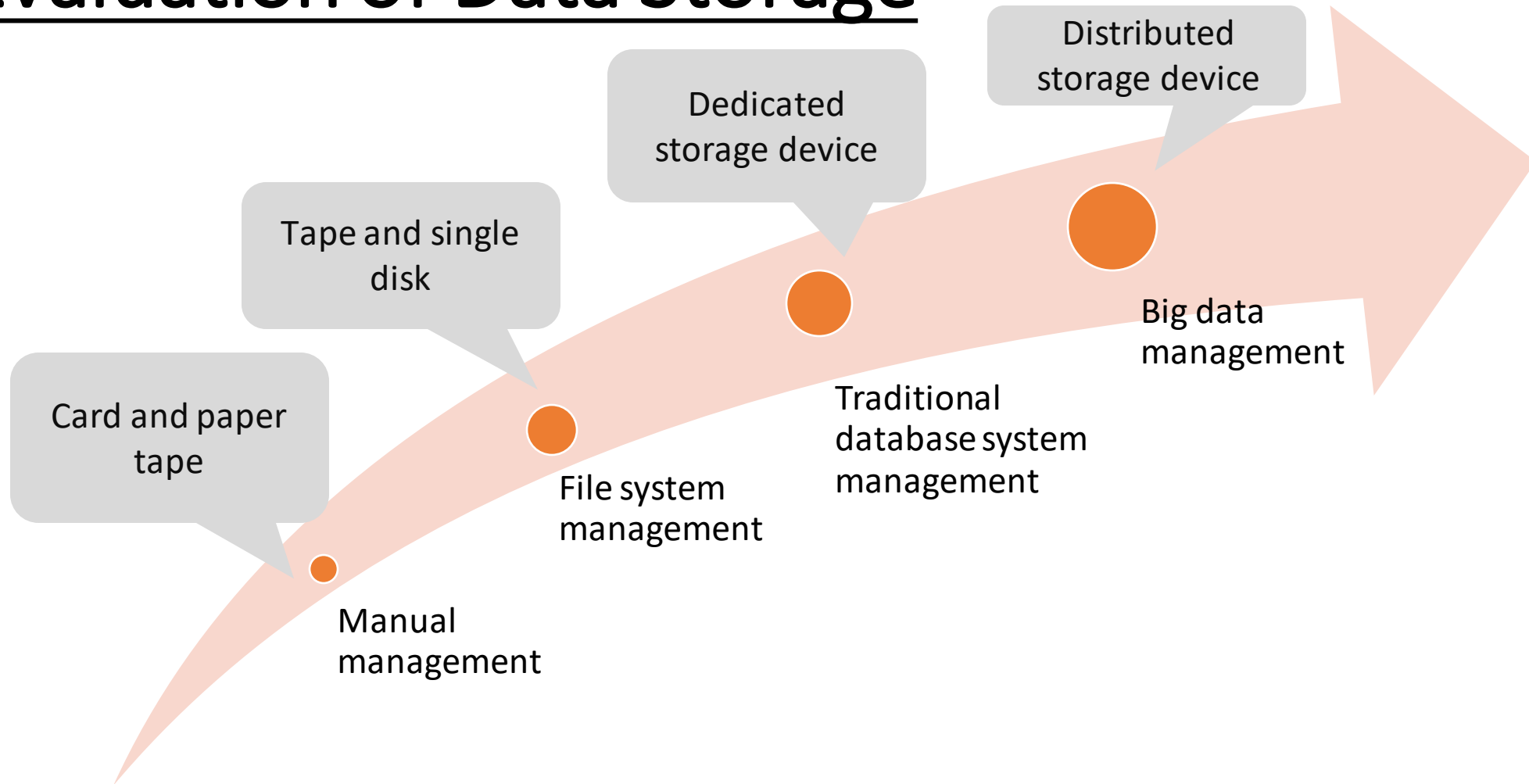
Data Storage Types



Data Storage Types

- Storage systems can be classified into internal and external storage systems based on the locations of storage devices and hosts.
- An internal storage system is directly connected to the host bus, and includes the high-speed cache and memory required for CPU computing, as well as disks and CD-ROM drives directly connected to the main boards of computers. Its capacity is generally small and hard to expand.
- An external storage system is classified into direct-attached storage (DAS) and fabric-attached storage (FAS) by connection mode.
- FAS is classified into two types: network-attached storage (NAS) and storage area network (SAN) by transmission protocol.

Evaluation of Data Storage



Evaluation of Data Storage

- Data management is a process of effectively collecting, storing, processing, and applying data using computer hardware and software to make full use of data. Data organization is the key to effective data management.
- Data management technology is used to classify, organize, encode, input, store, retrieve, maintain, and output data. The evolution of data storage devices and computer application systems promotes the development of databases and data management technologies. Data management in a computer system goes through four phases: manual management, file system management, traditional database system management, and big data management.

Contents

1. Data and Information

2. Data Storage

3. Development of Storage Technologies

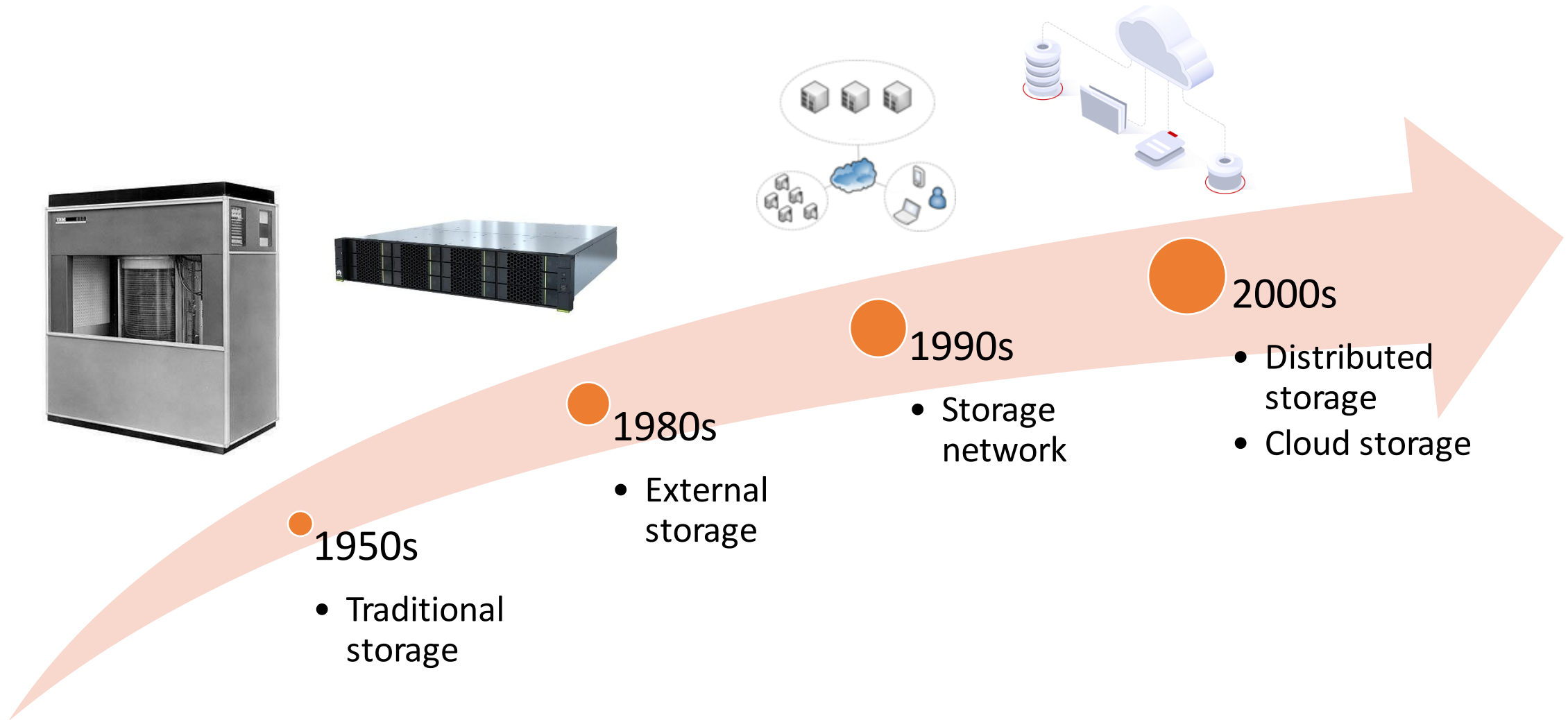
- Storage Architecture

- Storage Media

- Interface Protocols

4. Development Trend of Storage Products

History of Storage Architecture Development



History of Storage Architecture Development

- The storage architecture we use today derives from traditional storage, external storage, and storage networks, and has developed into distributed and cloud storage.
- Traditional storage is composed of disks. In 1956, IBM invented the world's first hard disk drive, which used 50 x 24-inch platters with a capacity of only 5 MB. It was as big as two refrigerators and weighed more than a ton. It was used in the industrial field at that time and was independent of the mainframe.
- External storage is also called direct attached storage. Its earliest form is Just a Bunch of Disks (JBOD), which simply combines disks and is represented to hosts as a bundle of independent disks. It only increases the capacity and cannot ensure data security.
- SAN is a typical storage network. It first emerged as FC SAN using the Fiber Channel network to transmit data, and later supported IP SAN.
- Distributed storage uses general-purpose server hardware to build storage resource pools and is applicable to cloud computing scenarios.