
Xử lý và lưu trữ dữ liệu

Môn học: Nhập môn khoa học máy tính

Mục tiêu



- Giới thiệu về dữ liệu
- Xử lý dữ liệu các hệ cơ sở
- Đơn vị thông tin

Giới thiệu về dữ liệu

Dữ liệu là gì?

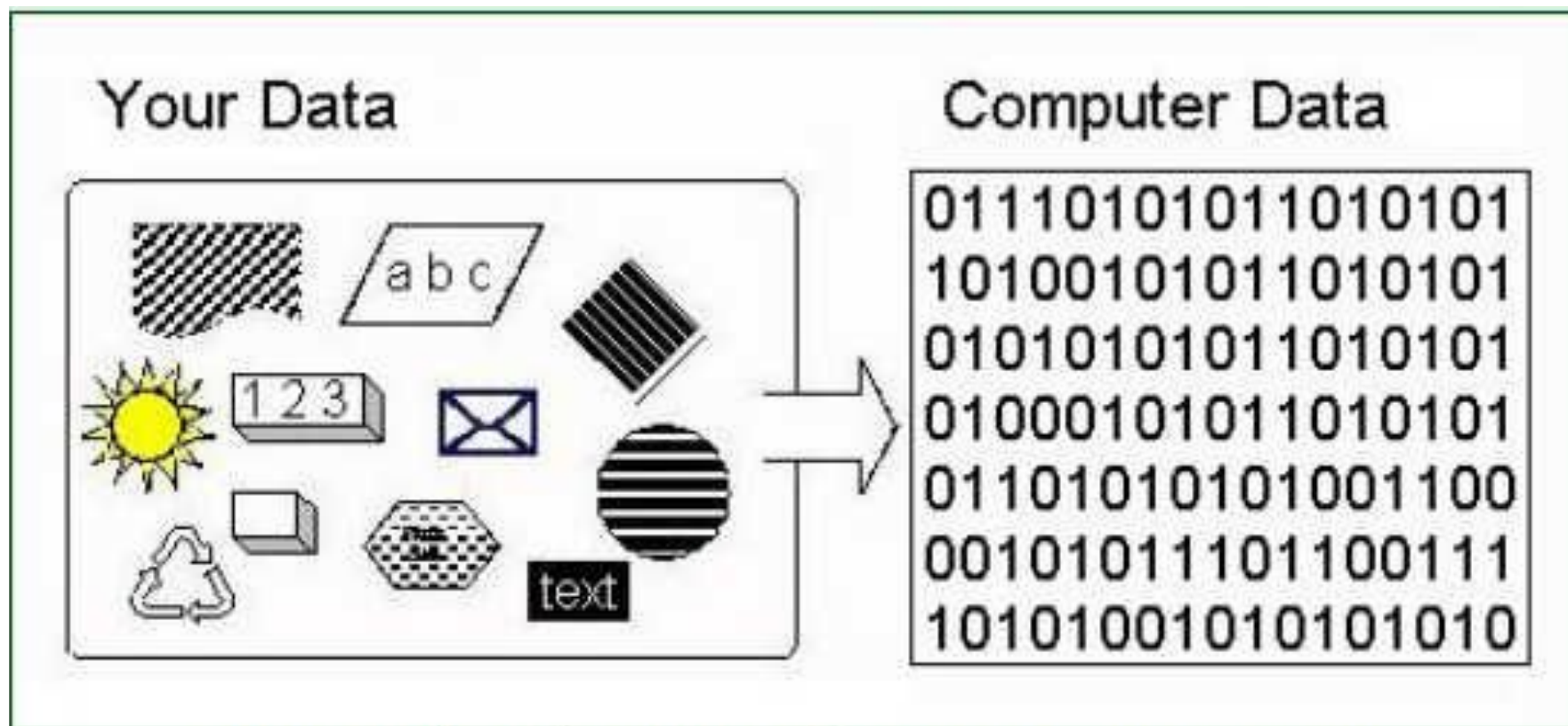
CPU, ROM, RAM

Thiết bị lưu trữ

Dữ liệu là gì



Dữ liệu là một tập hợp các dữ kiện, chẳng hạn như số, từ, hình ảnh, nhằm đo lường, quan sát hoặc chỉ là mô tả về sự vật



Bit & Byte



Bit là đơn vị nhỏ nhất dùng để biểu diễn thông tin trong máy tính.

Byte là đơn vị cực kỳ quan trọng, vì các tổ hợp đầy đủ 8 bit khác nhau biểu diễn tất cả các ký tự trên bàn phím, bao gồm các từ, các số, các dấu câu và các ký hiệu khác

Đơn vị	Viết tắt	Sức chứa - Tương đương
Bit	b	1 hoặc 0
Byte	B	1 Byte = 8 bits
Kilobyte	KB	1 Kilobyte = 1024 byte
Megabyte	MB	1 Megabyte = 1024 kilobyte
Gigabyte	GB	1 Gigabyte = 1024 megabyte
Terabyte	TB	1 Terabyte = 1024 gigabyte
Petabyte	PB	1 Petabyte = 1024 terabyte
Exabyte	EB	1 Exabyte = 1024 petabyte
Zettabyte	ZB	1 Zettabyte = 1024 exabyte
Yottabyte	YB	1 Yottabyte = 1024 zettabyte

Mã văn bản

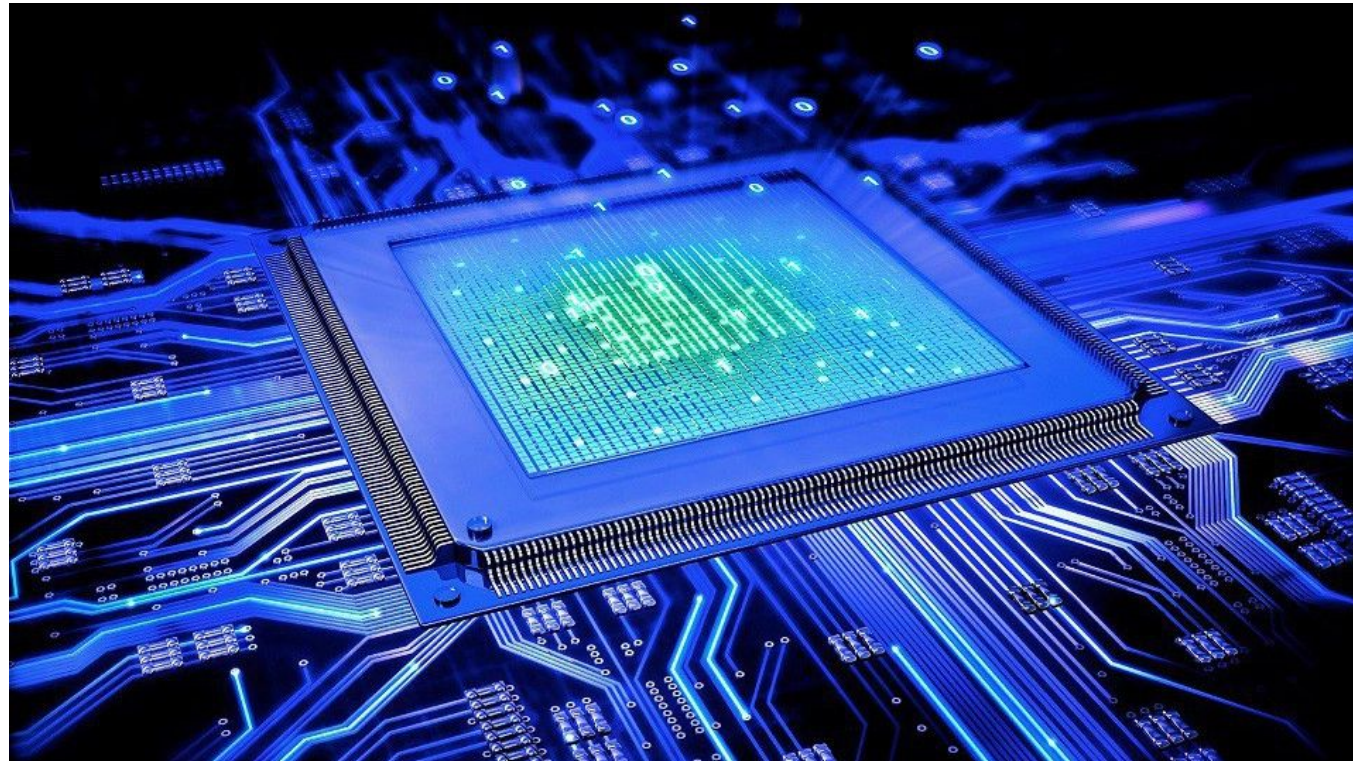


- Bảng mã tiêu chuẩn ASCII (**A**merican **S**tandard for **I**nformation **I**nterchange) dùng 7 bit để mã hóa 1 ký tự, cung cấp 128 sắp xếp khác nhau
- Bảng mã EBCDIC (**E**xtended **B**inary **C**oded **D**ecimal **I**nterchange **C**ode) dùng 8 bit có thể biểu diễn 256 ký tự
- Bảng mã Unicode dùng 2 byte = 16 bit có thể biểu diễn hơn 65536 các ký tự và ký hiệu khác nhau.

CPU



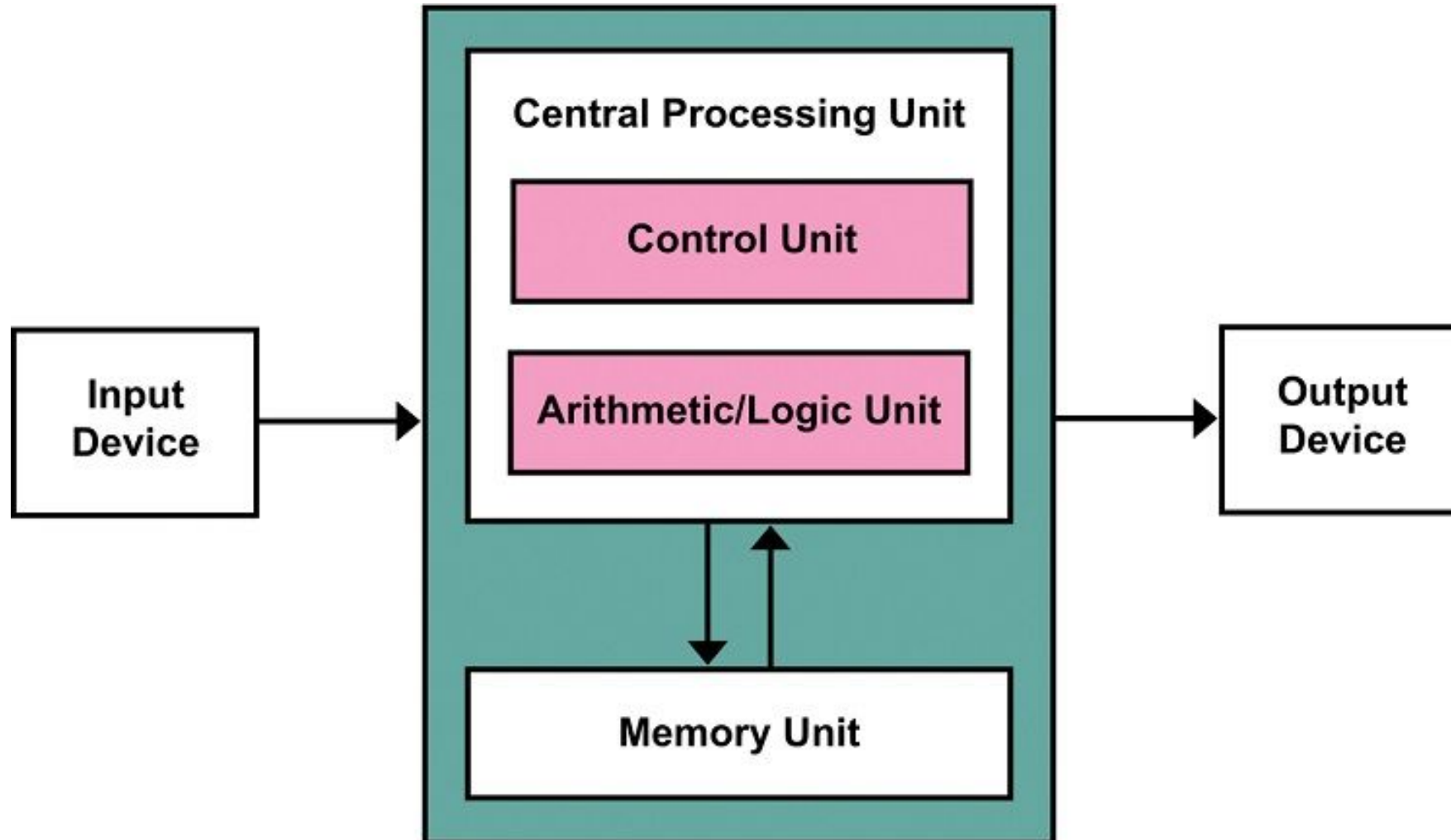
- Central Processing Unit - bộ xử lý trung tâm hay bộ não của máy tính là mạch điện tử thực hiện các câu lệnh của chương trình máy tính bằng các phép tính số học, logic, so sánh.



CPU gồm 2 bộ phận cơ bản:

- **Bộ điều khiển** (Control Unit – CU) có trách nhiệm tải và diễn giải các chỉ thị riêng lẻ bao gồm chương trình máy tính
- **Bộ số học và logic** (Arithmetic Logic Unit – ALU) là một mạch điện tử có nhiệm vụ thực hiện các phép tính số học và logic. Bộ logic số học sẽ nhận các lệnh điều khiển từ bộ điều khiển và tiến hành xử lý tín hiệu, thực hiện các phép tính số học.

CPU



Bộ nhớ



Bộ nhớ là một phần căn bản của một máy tính nơi lưu trữ dữ liệu cho công việc xử lý. Dữ liệu được xử lý bên trong CPU nhưng nó cần không gian hàng triệu byte để có thể nhanh chóng đọc/ghi các chương trình và dữ liệu trong khi chúng được sử dụng.

Chia làm 2 loại:

- Bộ nhớ lâu dài (không mất dữ liệu)
- Bộ nhớ tạm thời (có thể mất dữ liệu)

Read-Only Memory là một bộ nhớ lưu trữ lâu dài. Bởi vì mục đích của loại bộ nhớ này được sử dụng để lưu trữ dữ liệu quan trọng chỉ đọc, không được ghi, nên nó được gọi là bộ nhớ chỉ đọc

Các loại ROM:

- PROM chỉ có thể lập trình 1 lần
- EPROM (Erasable Programmable ROM) cho phép người dùng xóa và lập trình nhiều lần trên ROM
- EEPROM (Electrically Erasable Programmable ROM) có thể được lập trình thông qua các xung điện đặc biệt

RAM



Random Access Memory - Bộ nhớ truy cập ngẫu nhiên, chịu trách nhiệm lưu giữ dữ liệu và chương trình trong khi chúng đang sử dụng. Đây là bộ nhớ tạm thời vì dữ liệu trong RAM sẽ mất nếu máy tính bị tắt

Các loại RAM:

- DRAM
- SDRAM (Synchronous Dynamic RAM)
- DDR SDRAM (Double Data Rate Synchronous Dynamic RAM)
- SRAM

Thiết bị lưu trữ

Thiết bị lưu trữ từ tính:

- Đĩa mềm
- Ổ cứng

Thiết bị lưu trữ quang học:

- Đĩa CD-ROM
- DVD-R

Thiết bị nhớ flash:

- Thẻ nhớ
- USB
- NVMe



Xử lý dữ liệu các hệ cơ số

Giới thiệu các hệ cơ số

Chuyển đổi giữa các hệ cơ số

Các hệ cơ số



Hệ cơ số nhị phân (Binary system - hệ đếm cơ số 2) là hệ đếm dùng 2 ký tự (0,1) để biểu đạt giá trị số. Thường được dùng để biểu đạt giá trị hiệu điện thế tương ứng:

- Hiệu điện thế cao hoặc ý nghĩa có sẽ mang giá trị 1
- Hiệu điện thế thấp hoặc ý nghĩa không sẽ mang giá trị 0



1011001_2

Các hệ cơ số



- Hệ bát phân (Octal system - hệ đếm cơ số 8) dùng 8 ký tự (0,1,2,3,4,5,6,7) để biểu đạt giá trị số.
- Hệ thập phân (Decimal system - hệ đếm cơ số 10) dùng 10 ký tự (0,1,2,3,4,5,6,7,8,9) để biểu diễn giá trị số.
- Hệ thập lục phân (Hexadecimal system) dùng 16 ký tự (0,1,2,3,4,5,6,7,8,9,A,B,C,D,E,F) để biểu diễn giá trị số

Chuyển đổi giữa các hệ cơ số

Chuyển đổi giữa các hệ cơ số



Chuyển đổi số $(10110101)_2$ sang hệ cơ số 10

10110101

$$= 1 \times 2^7 + 0 \times 2^6 + 1 \times 2^5 + 1 \times 2^4 + 0 \times 2^3 + 1 \times 2^2 + 0 \times 2^1 + 1 \times 2^0$$

$$= 181$$

Chuyển đổi giữa các hệ cơ số



Chuyển đổi số $(13)_{10}$ sang hệ cơ số 2

1. Chia số cho 2.
2. Lấy thương số nguyên cho lần lặp tiếp theo.
3. Lấy phần dư cho chữ số nhị phân.
4. Lặp lại các bước cho đến khi thương số bằng 0

Chia cho 2	Thương số	Phần dư
13/2	6	1
6/2	3	0
3/2	1	1
1/2	0	1

$$\Rightarrow (13)_{10} = 1101$$

Chuyển đổi giữa các hệ cơ số



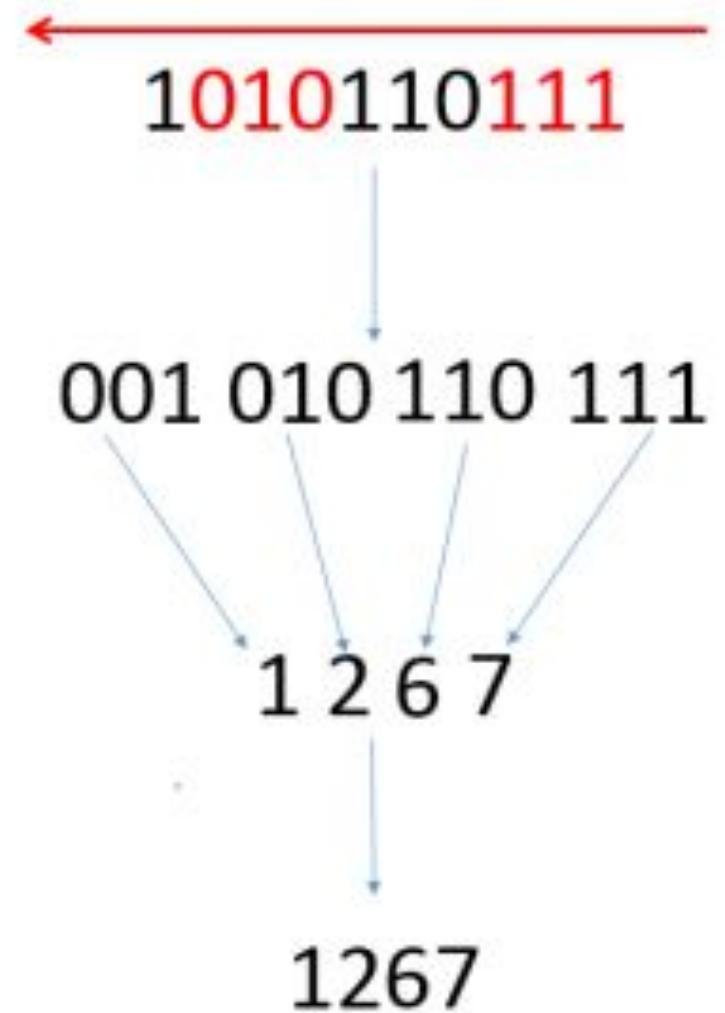
Thập phân (DEC)	Nhị phân (BIN)	Bát phân (OCT)	Thập lục phân (HEX)
0	0000	0	0
1	0001	1	1
2	0010	2	2
3	0011	3	3
4	0100	4	4
5	0101	5	5
6	0110	6	6
7	0111	7	7
8	1000		8
9	1001		9
10	1010		A
11	1011		B
12	1100		C
13	1101		D
14	1110		E
15	1111		F

Chuyển đổi giữa các hệ cơ số



Chuyển đổi dãy số nhị phân
1010110111 sang bát phân

Nhóm lần lượt các bit nhị phân từ
phải qua trái thành các nhóm 3
bit, sau đó dựa vào bảng chuyển
đổi để chuyển sang dãy số bát
phân

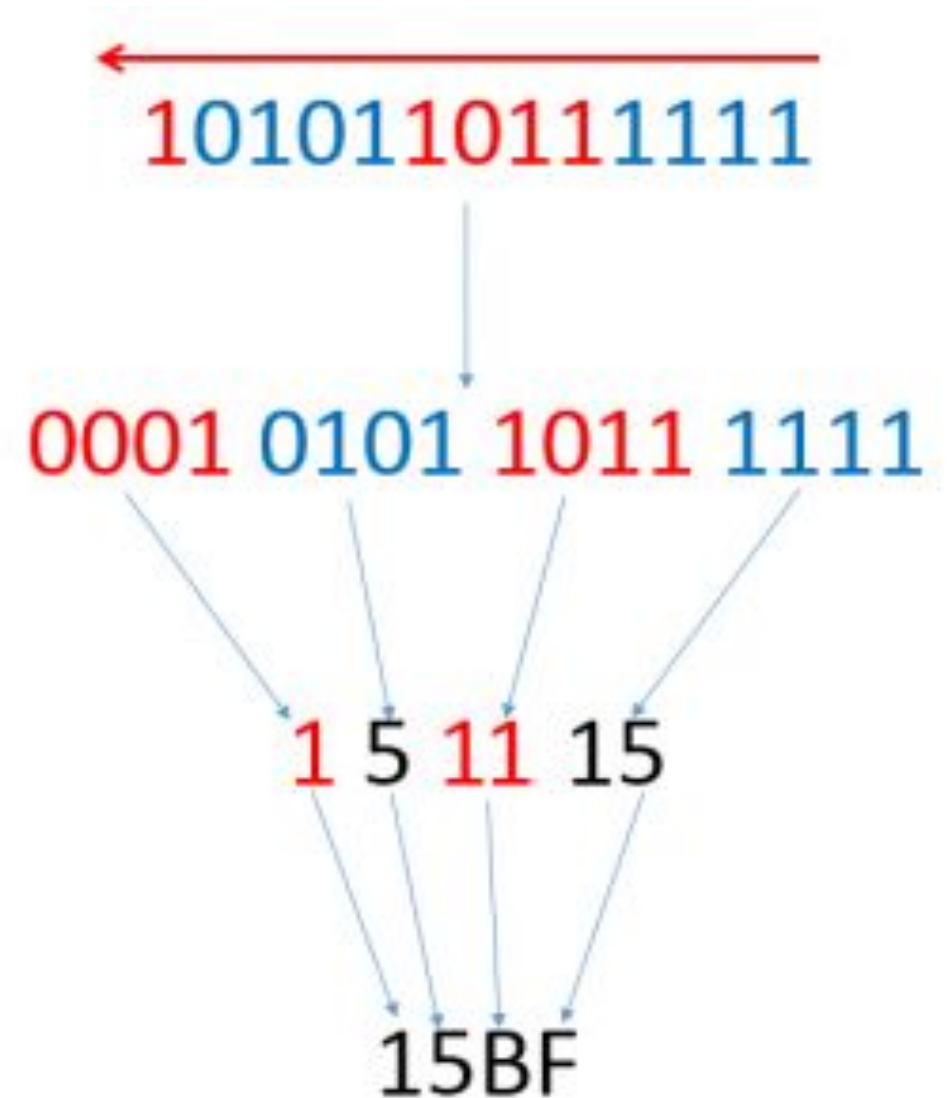


Chuyển đổi giữa các hệ cơ số



Chuyển đổi dãy số nhị phân
101011011111 sang thập lục
phân

Nhóm lần lượt các bit nhị phân từ
phải qua trái thành các nhóm 4
bit, sau đó dựa vào bảng chuyển
đổi để chuyển sang dãy số thập
lục phân

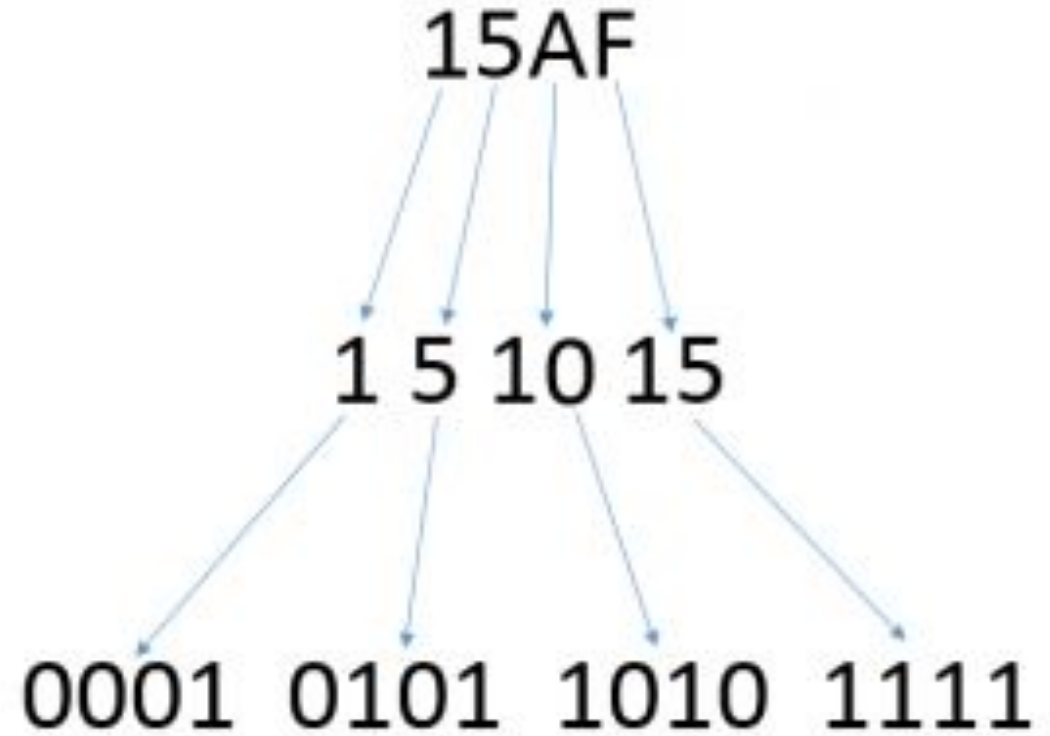
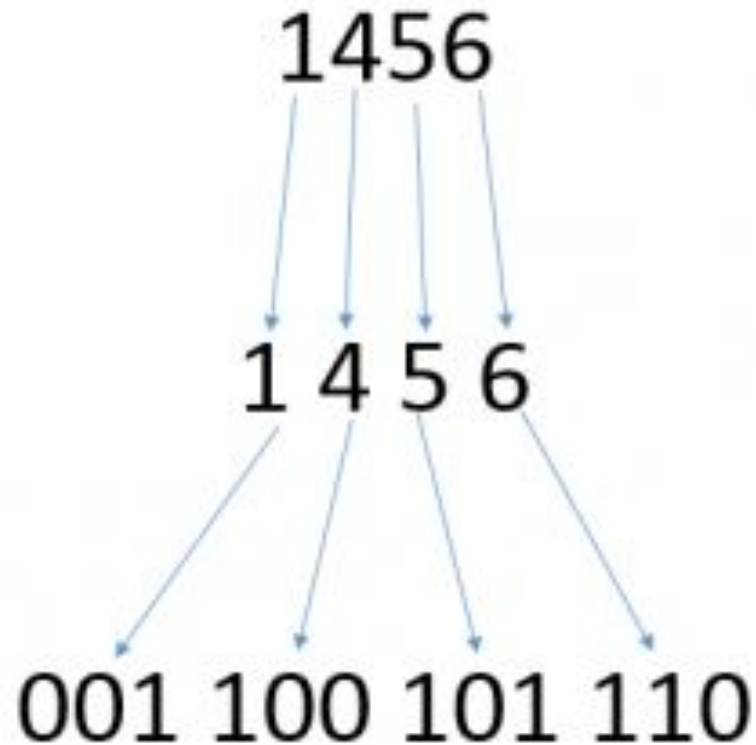


Chuyển đổi giữa các hệ cơ số



$$(1456)_8 = (1100101110)_2$$

$$(15AF)_{16} = (1010110101111)_2$$





Đơn vị thông tin

- Hầu hết các loại máy tính dùng các hệ số nhị phân đã mã hóa để biểu diễn ký tự, ký hiệu đặc biệt, số thập phân...
- 7 ký số được dùng để biểu diễn cho 128 ký tự riêng biệt.
- Việc mã hoá đã được chuẩn hoá nhằm đảm bảo khả năng trao đổi dữ liệu giữa các máy tính.

Bảng mã ASCII



- Một tiêu chuẩn phổ biến là American Standard Code for Information Interchange (ASCII)
- ASCII sử dụng 7 bit cho mỗi ký tự có thể tạo ra 128 (2^7) tổ hợp sắp xếp khác nhau.
- Những mã riêng biệt dùng để thể hiện thông tin khác như cuối một tập tin, đầu trang,... Các mã này được gọi là ký tự điều khiển không in được.
- Mã ASCII được dùng để thể hiện dữ liệu bên trong các máy tính cá nhân.

Bảng mã EBCDIC



- Extended Binary Coded Decimal Interchange Code (EBCDIC)
- EBCDIC dùng 8 bit cho mỗi ký tự.
- Như vậy EBCDIC có thể biểu diễn được 256 ký tự.
- Mã EBCDIC được dùng trong các hệ thống Mainframe IBM và những hệ máy tương đương khác.
- Tín hiệu điện tử có khả năng chuyển đổi ký tự từ dạng ASCII sang EBCDIC và ngược lại.
- Chúng ta cũng có thể có được các giá trị chuyển đổi này thông qua các chương trình máy tính.



Tóm tắt bài học

Giới thiệu về dữ liệu:

- Dữ liệu là gì?
- CPU, ROM, RAM
- Thiết bị lưu trữ

Xử lý dữ liệu các hệ cơ số:

- Giới thiệu các hệ cơ số
- Chuyển đổi qua lại giữa các hệ cơ số

Đơn vị thông tin: Bảng mã ASCII, EBCDIC

Hướng dẫn

Hướng dẫn làm bài thực hành và bài tập

Chuẩn bị bài tiếp