



COMPUTER ARCHITECTURE AND SOFTWARE EXECUTION PROCESS MEMORY MANAGEMENT

Bachelor in Artificial Intelligence, Data and Management Sciences

m CentraleSupelec and ESSEC Business School - 2024/2025



OUTLINE

- The main memory
- Memory allocation strategies
- ▶ The Virtual memory

Back to the begin - Back to the outline



OUTLINE

- The main memory
- Memory allocation strategies
- ▶ The Virtual memory

Back to the begin - Back to the outline





For the operating system



For the operating system

 When a machine is started, the OS is the first program loaded into memory.



For the operating system

- When a machine is started, the OS is the first program loaded into memory.
- The OS needs a memory space for:
 - the code of its Core
 - the interruptions table
 - the processes table
 - data structures (PCBs and others)
 - IIII



For processes (running programs)



For processes (running programs)

 When a process is created, the OS creates a PCB and allocates memory for the process.



For processes (running programs)

- When a process is created, the OS creates a PCB and allocates memory for the process.
- For security reasons, each process must use a separate memory area (an address space).

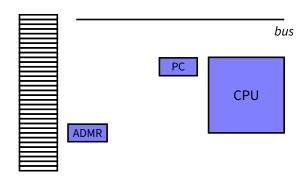


For processes (running programs)

- When a process is created, the OS creates a PCB and allocates memory for the process.
- For security reasons, each process must use a separate memory area (an address space).
 - which mechanism for allocating this space?
 - how to ensure the protection of this area?
 - how to ensure the transparency of this space?

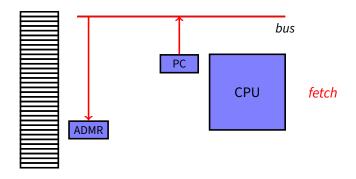


- instructions → PC register on the processor
- data → ADMR register on the processor



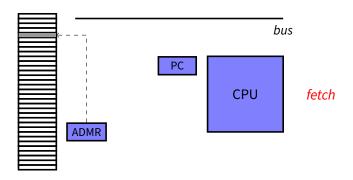


- instructions → PC register on the processor
- data → ADMR register on the processor



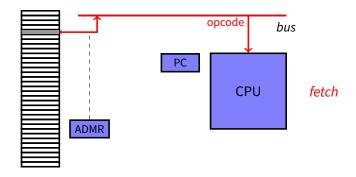


- instructions → PC register on the processor
- data → ADMR register on the processor



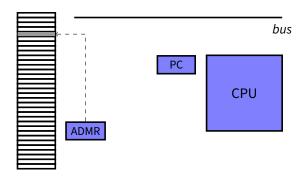


- instructions → PC register on the processor
- data → ADMR register on the processor



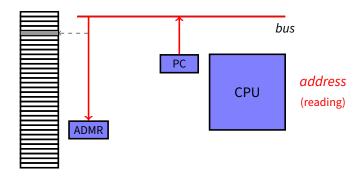


- instructions → PC register on the processor
- data → ADMR register on the processor



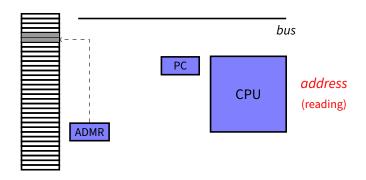


- instructions → PC register on the processor
- data → ADMR register on the processor



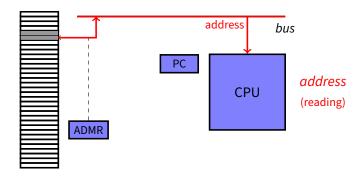


- instructions → PC register on the processor
- data → ADMR register on the processor



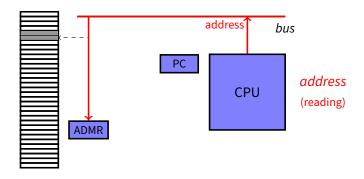


- instructions → PC register on the processor
- data → ADMR register on the processor



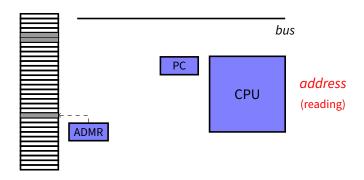


- instructions → PC register on the processor
- data → ADMR register on the processor



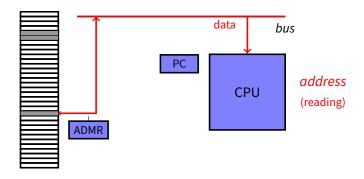


- instructions → PC register on the processor
- data → ADMR register on the processor



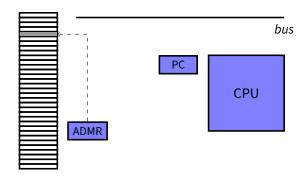


- instructions → PC register on the processor
- data → ADMR register on the processor



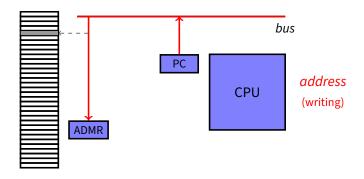


- instructions → PC register on the processor
- data → ADMR register on the processor



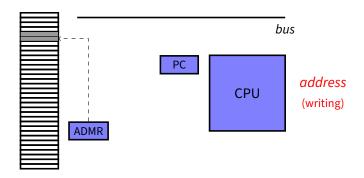


- instructions → PC register on the processor
- data → ADMR register on the processor



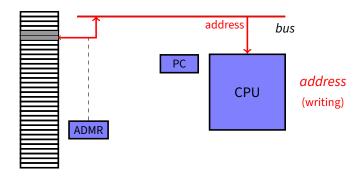


- instructions → PC register on the processor
- data → ADMR register on the processor



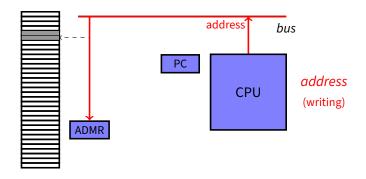


- instructions → PC register on the processor
- data → ADMR register on the processor



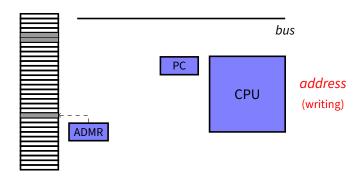


- instructions → PC register on the processor
- data → ADMR register on the processor



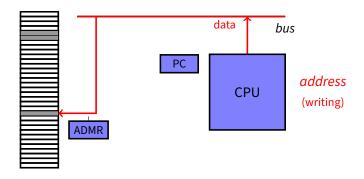


- instructions → PC register on the processor
- data → ADMR register on the processor





- instructions → PC register on the processor
- data → ADMR register on the processor







- Each memory word is associated with an address
 - the box number



- Each memory word is associated with an address
 - the box number
- This address is obtained from an instruction
 - ightharpoonup the address is in binary with n bits



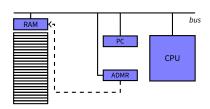
- Each memory word is associated with an address
 - the box number
- This address is obtained from an instruction
 - the address is in **binary** with n bits
- So there are 2^n different addresses
 - \rightarrow 2ⁿ boxes from 0000 ... 0 to 1111 ... 1



- Each memory word is associated with an address
 - the box number
- This address is obtained from an instruction
 - \implies the address is in binary with n bits
- So there are 2^n different addresses
 - $ightharpoonup 2^n$ boxes from $0000 \dots 0$ to $1111 \dots 1$
- Example : 32 bits $\rightarrow 2^{32} \approx 4$ GB



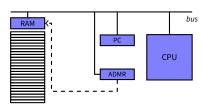
HOW MEMORY WORKS





HOW MEMORY WORKS

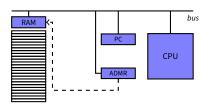
 The CPU will fetch instructions from memory by using their address (fetch);





HOW MEMORY WORKS

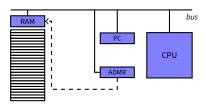
- The CPU will fetch instructions from memory by using their address (fetch);
- The CPU will retrieve data from variables in memory by using their address;





HOW MEMORY WORKS

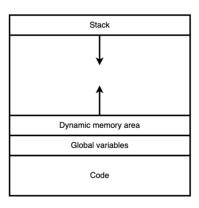
- The CPU will fetch instructions from memory by using their address (fetch);
- The CPU will retrieve data from variables in memory by using their address;
- The CPU writes on variables at a given address in memory.





ADDRESS SPACE USAGE

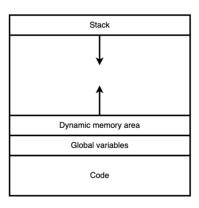
• What does a process's memory space contain?





ADDRESS SPACE USAGE

- What does a process's memory space contain?
 - Code (known size)
 - Global variables (known size)
 - Stack (unknown size)
 - Dynamic memory area (unknown size)
 - **.**.







A program (code + data) is loaded from the disk to the memory ...



A program (code + data) is loaded from the disk to the memory ... it is placed at a given location in the memory



A program (code + data) is loaded from the disk to the memory ... it is placed at a given location in the memory

Question 🗳

what are the addresses of the variables in memory?





```
1 int a = 3;
2 a = a + 2;
```



```
1 int a = 3;

2 a = a + 2;

1 @a: memval 3

2 mov eax, a

3 mov ebx, 2

4 add ecx, eax, ebx

5 mov a, ecx
```



```
1 int a = 3;
2 a = a + 2;
2 mov eax, a
3 mov ebx, 2
4 add ecx, eax, ebx
5 mov a, ecx
1 2850: mov eax, 2B1E
2 2852: mov ebx, #0002
3 2854: add ecx, eax, ebx
4 2855: mov 2B1E, ecx
5 ...
6 281E: 0003
```



symbolic addresses vs memory addresses

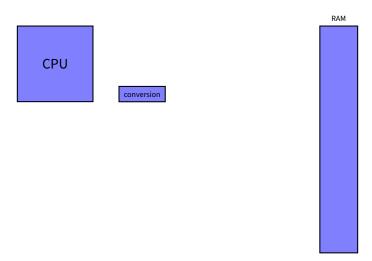
```
1 int a = 3;
2 a = a + 2;
2 mov eax, a
3 mov ebx, 2
4 add ecx, eax, ebx
5 mov a, ecx
1 2B50: mov eax, 2B1E
2 2B52: mov eax, 2B1E
2 2B52: mov ebx, #0002
3 2B54: add ecx, eax, ebx
4 2B55: mov 2B1E, ecx
5 ...
6 2B1E: 0003
```

Link editing

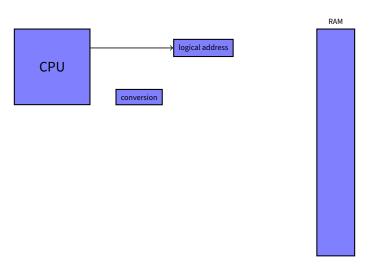
When creating processes, the **OS** instantiates the program.

transform names of variables into addresses.

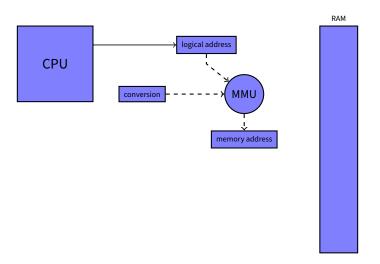




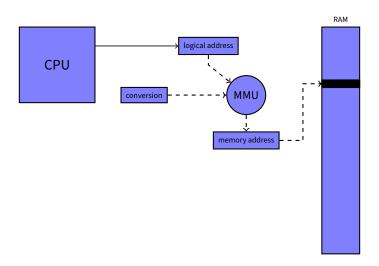














OUTLINE

- The main memory
- Memory allocation strategies
- ▶ The Virtual memory

Back to the begin - Back to the outline





• A strategy for allocating and freeing memory must be chosen based on the needs of the processes.



- A strategy for allocating and freeing memory must be chosen based on the needs of the processes.
- Two possible strategies :



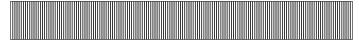
- A strategy for allocating and freeing memory must be chosen based on the needs of the processes.
- Two possible strategies :
 - 1. Contiguous allocation of memory slots (by partition)



- A strategy for allocating and freeing memory must be chosen based on the needs of the processes.
- Two possible strategies :
 - 1. Contiguous allocation of memory slots (by partition)
 - 2. Non-contiguous allocation (by pagination)

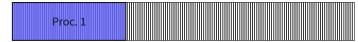


Process constitutes a single, non-decomposable block.



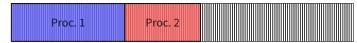


Process constitutes a single, non-decomposable block.



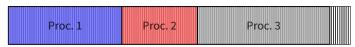


Process constitutes a single, non-decomposable block.



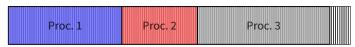


Process constitutes a single, non-decomposable block.





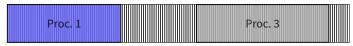
Process constitutes a single, non-decomposable block.





Process constitutes a single, non-decomposable block.

RAM

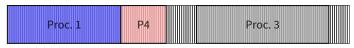


X holes appear → **fragmentation**



Process constitutes a single, non-decomposable block.

RAM

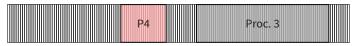


x holes appear → **fragmentation**



Process constitutes a single, non-decomposable block.

RAM

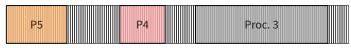


X holes appear → **fragmentation**



Process constitutes a single, non-decomposable block.

RAM



x holes appear → **fragmentation**



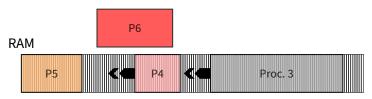
Process constitutes a single, non-decomposable block.



- **X** holes appear → **fragmentation**
- ✗ bigs processes cannot fit in



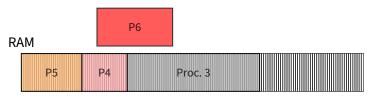
Process constitutes a single, non-decomposable block.



- **X** holes appear → **fragmentation**
- x bigs processes cannot fit in \rightarrow **defragment**



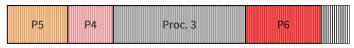
Process constitutes a single, non-decomposable block.



- **X** holes appear → **fragmentation**
- \times bigs processes cannot fit in \rightarrow defragment



Process constitutes a single, non-decomposable block.

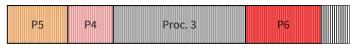


- **X** holes appear → **fragmentation**
- x bigs processes cannot fit in \rightarrow **defragment**



Process constitutes a single, non-decomposable block.

RAM

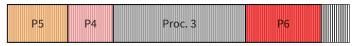


- **X** holes appear → **fragmentation**
- \times bigs processes cannot fit in \rightarrow defragment
- reduce **fragmentation**



Process constitutes a single, non-decomposable block.

RAM



- **X** holes appear → **fragmentation**
- \times bigs processes cannot fit in \rightarrow defragment
- reduce fragmentation
- reduce **defragmentation** operations





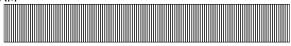
- Pros
 - ✓ Material simplicity
 - ✓ Transparency for programs
 - ✔ Checking the validity of addresses



- Pros
 - ✓ Material simplicity
 - ✓ Transparency for programs
 - ✓ Checking the validity of addresses
- Cons
 - **X** Fragmentation
 - ✗ Fixed size of memory spaces



RAM

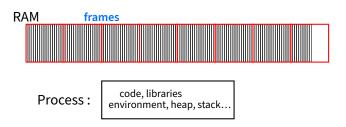


Process:

code, libraries environment, heap, stack…



• Divide the physical memory (Main memory) into blocks (frames).



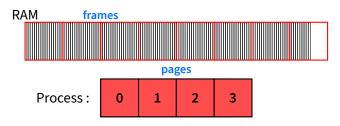


- Divide the physical memory (Main memory) into blocks (frames).
- Divide the address space used by a process (logical space) into pages.



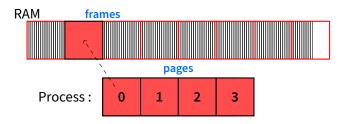


- Divide the physical memory (Main memory) into blocks (frames).
- Divide the address space used by a process (logical space) into pages.
 - each page has the same size as a frame



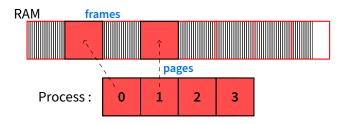


- Divide the physical memory (Main memory) into blocks (frames).
- Divide the address space used by a process (logical space) into pages.
 - each page has the same size as a frame
- Place the pages in the frames



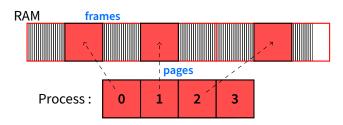


- Divide the physical memory (Main memory) into blocks (frames).
- Divide the address space used by a process (logical space) into pages.
 - each page has the same size as a frame
- Place the pages in the frames



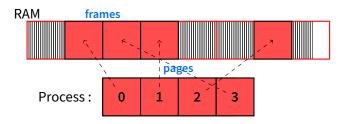


- Divide the physical memory (Main memory) into blocks (frames).
- Divide the address space used by a process (logical space) into pages.
 - each page has the same size as a frame
- Place the pages in the frames





- Divide the physical memory (Main memory) into blocks (frames).
- Divide the address space used by a process (logical space) into pages.
 - each page has the same size as a frame
- Place the pages in the frames





- Memory allocation:
 - a process is in *disjoint* areas
 - no need to defragment



- Memory allocation :
 - a process is in disjoint areas
 - no need to defragment
- Adaptation :
 - ightharpoonup need more memory ightharpoonup add pages
 - no need to re-allocate it entirely



- Memory allocation :
 - a process is in disjoint areas
 - no need to defragment
- Adaptation:
 - \longrightarrow need more memory \rightarrow *add* pages
 - no need to re-allocate it entirely
- Virtual memory: load only the pages that the process needs.

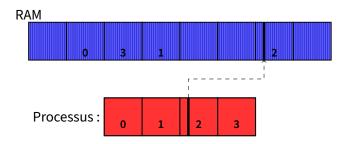




• Determine the physical address from the logical address



• Determine the physical address from the logical address





- Determine the physical address from the logical address
- Logical address
 - Page number (*n* bits) + offset (*m* bits)



- Determine the physical address from the logical address
- Logical address
 - Page number (n bits) + offset (m bits)
- Each process maintains a list:
 - lacksquare page number ightarrow frame number

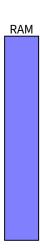


- Determine the physical address from the logical address
- Logical address
 - Page number (n bits) + offset (m bits)
- Each process maintains a list:
 - page number \rightarrow frame number
 - this is the pages table

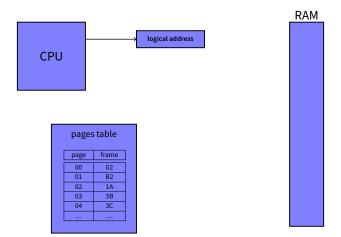




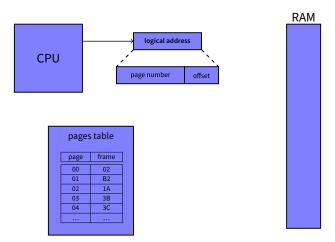




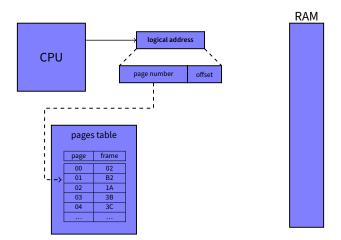




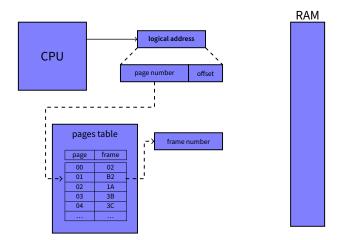




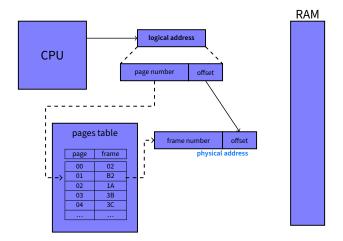




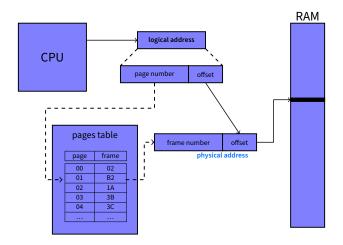














OUTLINE

- The main memory
- Memory allocation strategies
- The Virtual memory

Back to the begin - Back to the outline





• Number and size of processes :



- Number and size of processes :
 - between 200 and 500 parallel processes on a CPU



- Number and size of processes:
 - between 200 and 500 parallel processes on a CPU
 - big process : Eclipse = 250 MB, game data \geq 1 GB



- Number and size of processes :
 - between 200 and 500 parallel processes on a CPU
 - big process : Eclipse = 250 MB, game data \geq 1 GB
 - \mathbf{x} sum of process sizes \geq RAM capacity



- Number and size of processes :
 - between 200 and 500 parallel processes on a CPU
 - big process : Eclipse = 250 MB, game data ≥ 1 GB
 - \mathbf{x} sum of process sizes \geq RAM capacity
- Unused parts of code



- Number and size of processes :
 - between 200 and 500 parallel processes on a CPU
 - big process : Eclipse = 250 MB, game data ≥ 1 GB
 - **✗** sum of process sizes ≥ RAM capacity
- Unused parts of code
 - lacktriangledown error handling ightarrow rarely used



- Number and size of processes :
 - between 200 and 500 parallel processes on a CPU
 - big process : Eclipse = 250 MB, game data ≥ 1 GB
 - \mathbf{x} sum of process sizes \geq RAM capacity
- Unused parts of code
 - ullet error handling o rarely used
 - data (table, game, ...) → not all at once



- Number and size of processes :
 - between 200 and 500 parallel processes on a CPU
 - big process : Eclipse = 250 MB, game data ≥ 1 GB
 - **✗** sum of process sizes ≥ RAM capacity
- Unused parts of code
 - lacktriangledown error handling ightarrow rarely used
 - data (table, game, ...) \rightarrow not all at once
 - lacktriangle library ightarrow not all at once



- Number and size of processes :
 - between 200 and 500 parallel processes on a CPU
 - big process : Eclipse = 250 MB, game data ≥ 1 GB
 - \times sum of process sizes \geq RAM capacity
- Unused parts of code
 - lacktriangledown error handling ightarrow rarely used
 - data (table, game, ...) → not all at once
 - library \rightarrow not all at once
- load only useful pages!



- Number and size of processes :
 - between 200 and 500 parallel processes on a CPU
 - big process : Eclipse = 250 MB, game data ≥ 1 GB
 - ✗ sum of process sizes ≥ RAM capacity
- Unused parts of code
 - error handling → rarely used
 - data (table, game, ...) → not all at once
 - library \rightarrow not all at once
- load only useful pages!
- leave the rest on disk





• Extension of pagination mechanisms



- Extension of pagination mechanisms
 - a process organized into pages can represent spaces not present in RAM (eg disk) → virtual memory



- Extension of pagination mechanisms
 - a process organized into pages can represent spaces not present in RAM (eg disk) → virtual memory
 - pages are either in RAM or in auxiliary memory (swap) → RAM is a cache



- Extension of pagination mechanisms
 - a process organized into pages can represent spaces not present in RAM (eg disk) → virtual memory
 - pages are either in RAM or in auxiliary memory (swap) → RAM is a cache
- Each row of the page table contains:



- Extension of pagination mechanisms
 - a process organized into pages can represent spaces not present in RAM (eg disk) → virtual memory
 - pages are either in RAM or in auxiliary memory (swap) → RAM is a cache
- Each row of the page table contains:
 - a validity bit that indicates whether the page is in RAM

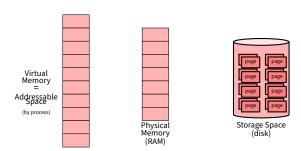


- Extension of pagination mechanisms
 - a process organized into pages can represent spaces not present in RAM (eg disk) → virtual memory
 - pages are either in RAM or in auxiliary memory (swap) → RAM is a cache
- Each row of the page table contains:
 - a validity bit that indicates whether the page is in RAM
 - the corresponding address in RAM (frame number)

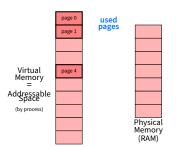


- Extension of pagination mechanisms
 - a process organized into pages can represent spaces not present in RAM (eg disk) → virtual memory
 - pages are either in RAM or in auxiliary memory (swap) → RAM is a cache
- Each row of the page table contains:
 - a validity bit that indicates whether the page is in RAM
 - the corresponding address in RAM (frame number)
 - otherwise information to find it on disk



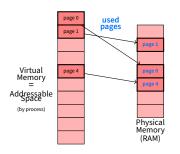






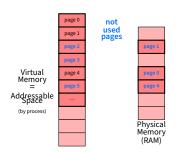






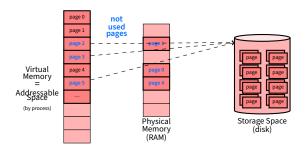
















Each process can address more space than it actually has in physical memory

• Access to a page not present in RAM:



- Access to a page not present in RAM:
 - raising a CPU exception → Page-Default



- Access to a page not present in RAM :
 - raising a CPU exception → Page-Default
 - current process blocked and loading the page into RAM



- Access to a page not present in RAM:
 - raising a CPU exception → Page-Default
 - current process blocked and loading the page into RAM
- Benefits



- Access to a page not present in RAM :
 - raising a CPU exception → Page-Default
 - current process blocked and loading the page into RAM
- Benefits
 - ✓ hide RAM size



Each process can address more space than it actually has in physical memory

- Access to a page not present in RAM :
 - raising a CPU exception → Page-Default
 - current process blocked and loading the page into RAM

Benefits

- ✓ hide RAM size
- ✓ possibility to put more processes in parallel



Each process can address more space than it actually has in physical memory

- Access to a page not present in RAM :
 - raising a CPU exception → Page-Default
 - current process blocked and loading the page into RAM

Benefits

- ✓ hide RAM size
- ✓ possibility to put more processes in parallel
- ✓ assign multiple virtual addresses to a physical address



Each process can address more space than it actually has in physical memory

- Access to a page not present in RAM :
 - raising a CPU exception → Page-Default
 - current process blocked and loading the page into RAM

Benefits

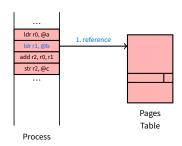
- ✓ hide RAM size
- ✓ possibility to put more processes in parallel
- ✓ assign multiple virtual addresses to a physical address
- ✓ pagination on demand



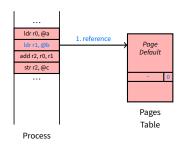


Process

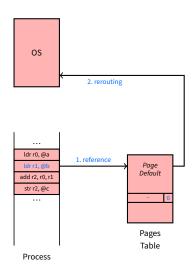




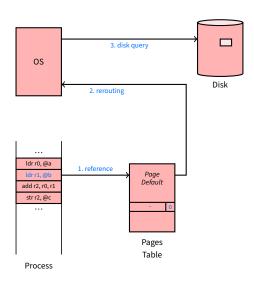




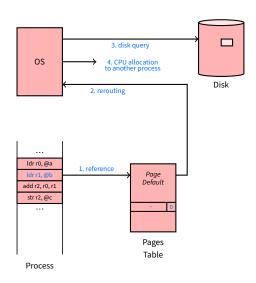




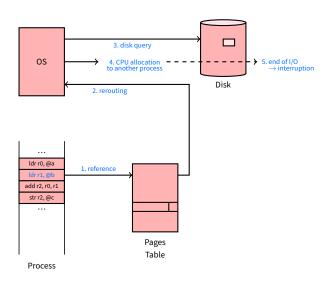




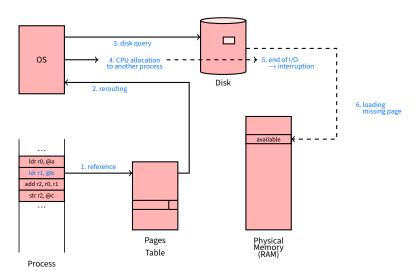




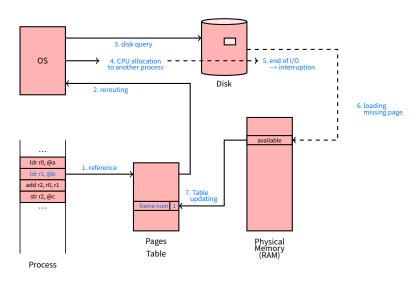




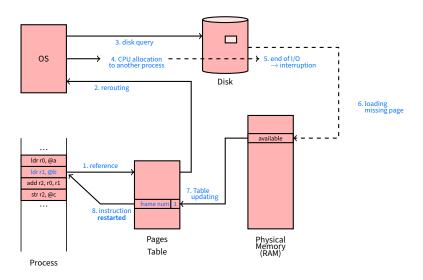














THANK YOU

Back to the begin - Back to the outline

