

UNIVERSITÉ GASTON BERGER DE SAINT-LOUIS



Introduction aux Systèmes Distribués

B. DIOP

MaDSI 1 - 2020



Sommaire



- Définitions & généralités
- Catégories de SDs
- Système multiprocesseurs
- Système parallèle multi-ordinateurs
- Taxinomie de Flynn

Dans la nature ...



Collection d'entités indépendantes coopérant pour exécuter une tâche globale qui ne peut pas être résolu individuellement



Parallélisme dans la vie courante

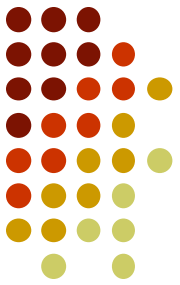


Le parallélisme au quotidien
de l'homme pour augmenter ses
performances ...



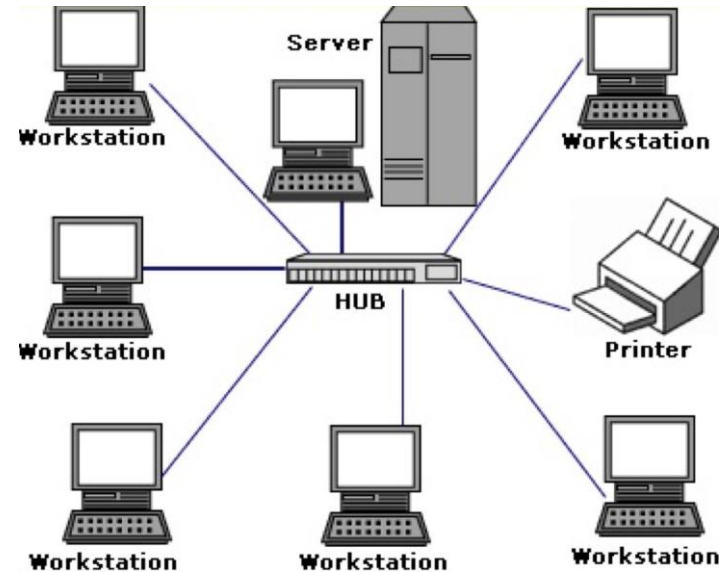
Contexte des SDs

Émergence des réseaux



1945

1985



Réseaux très haut débits

100 million à 10 milliards
bits/sec

Contexte des SDs

Histoire

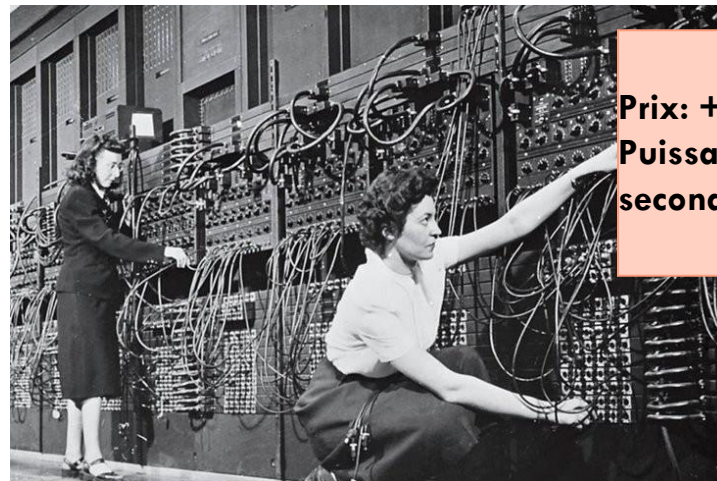
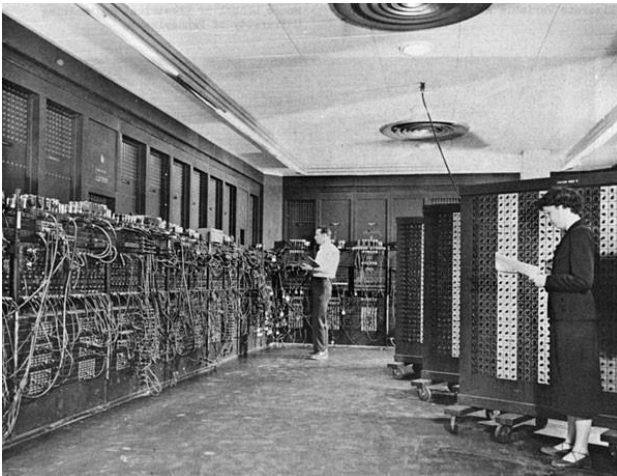


1945

1985

Ordinateurs grands et coûteux

Fonctionnement autonome des différentes machines



Prix: +1000 000 \$
Puissance: 1 instruction par
seconde

Contexte des SDs

Arrivée des microprocesseurs



1945

1985

Développement
des
microprocesseurs
8, 16, 32 et 64 bits

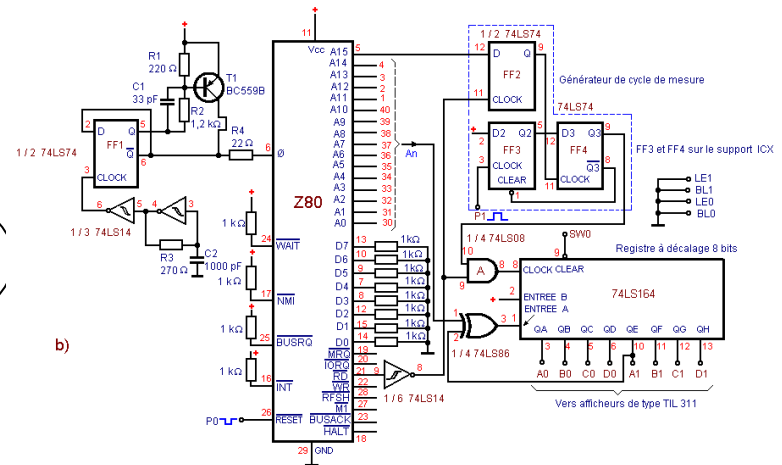
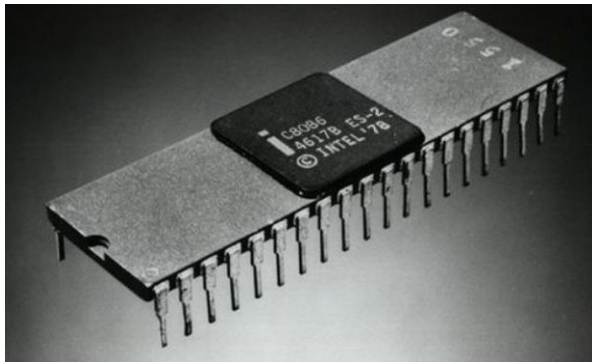


Fig. 18. - Schéma électrique du circuit de contrôle du microprocesseur Z80 par la méthode de la «signature analysis» (3ème expérience).



Prix: \$1000

Puissance: Milliers d'instructions
par seconde

Définition



Les systèmes distribués en Informatique se sont développés grâce à plusieurs facteurs :

- ✓ L'avancée de l'Internet et le développement des réseaux
- ✓ L'accessibilité des ordinateurs
- ✓ La réduction de coût des outils de calcul (processeurs, mémoire)
- ✓ L'augmentation de la bande passante des réseaux

SD selon Lamport



« Vous savez que vous en utilisez un SD lorsque le crash d'un ordinateur dont vous n'avez jamais entendu parler vous empêche de travailler. »

(Lamport)

SD selon Singhal-Shrivaratri



« Une collection d'ordinateurs qui ne partagent pas la mémoire commune ou une horloge physique commune, et qui communiquent par message passant sur un réseau de communication; et chaque ordinateur possède sa propre mémoire et exécute son propre système d'exploitation. Généralement, les ordinateurs sont semi-autonomes et sont faiblement couplés alors qu'ils coopèrent pour résoudre un problème collectivement. »

(Singhal-Shivaratri)

SD selon Tanenbaum



« Une collection d'ordinateurs indépendants qui apparaît aux utilisateurs du système comme un seul ordinateur cohérent. »

(Tanenbaum)

SD selon Goscinski

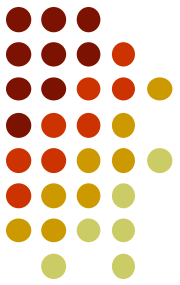


«Un terme qui décrit une large gamme d'ordinateurs, des systèmes faiblement couplés tels que les réseaux étendus aux systèmes fortement couplés tels que les réseaux locaux aux systèmes très fortement couplés tels que les systèmes multiprocesseurs.»

(Goscinski)

Systèmes distribués

définition synthétique



● Définition 1:

- Unités de traitements interconnectés, coopérant pour exécuter une tâche globale.**

● Définition 2:

- Ensemble composé d'éléments reliés par un système de communication, ayant des fonctions de traitement, de stockage et d'interaction avec le monde extérieur.**

Systèmes distribués

définition synthétique



- **L'utilisateur ne voit pas les différences entre les machines**
- **L'utilisateur ne sait pas quelle machine exécute un programme donné, où il l'exécute dans le système, ni quand.**
 - **Des ordinateurs/équipements indépendants coopèrent afin d'effectuer une tâche commune:**
 - **Exemple: Prédiction climatique, Simulation et calcul scientifiques,**

Caractéristiques clés



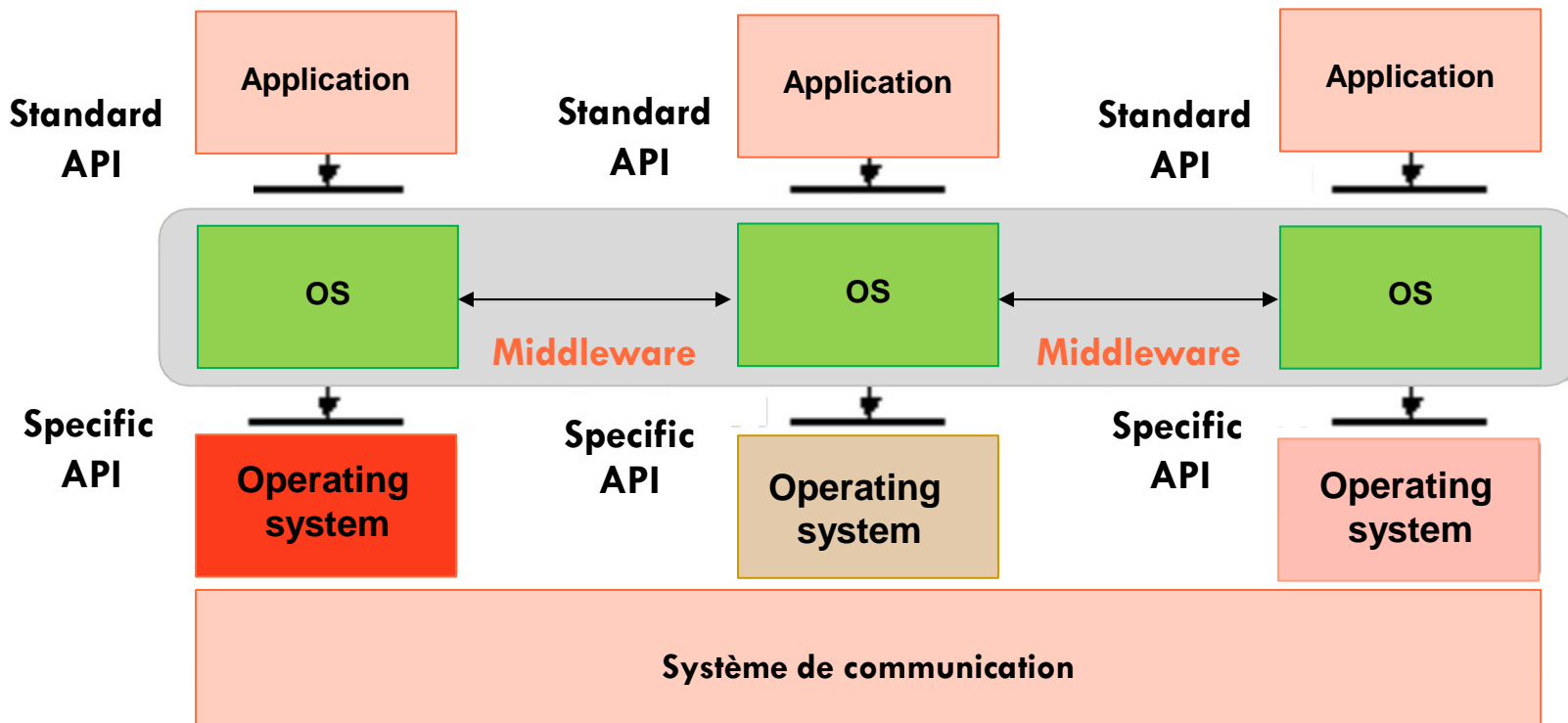
- ✓ Absence d'horloge physique globale (asynchronisme)
- ✓ Absence d'un état global
- ✓ Séparation géographique éventuelle
- ✓ Autonomie et hétérogénéité

Systemes distribués

Middleware



Couches logicielles permettant un fonctionnement global



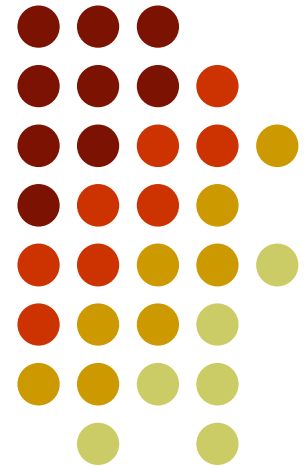
Systemes distribués



Objectifs

- **Accessibilité**
 - Partage de ressources, accès à distance, etc.
- **Transparence**
 - Format des données, location, concurrence, bugs,
- **Ouverture**
 - Règles et protocoles standards
- **Passage à l'échelle**
 - Capable de supporter des architectures complexes

Catégories de SD



Catégories de SD



- **systemes informatiques distribués**
- **systemes d'information distribués**
- **systemes embarqués**

Catégories de SD

Systèmes informatiques distribués



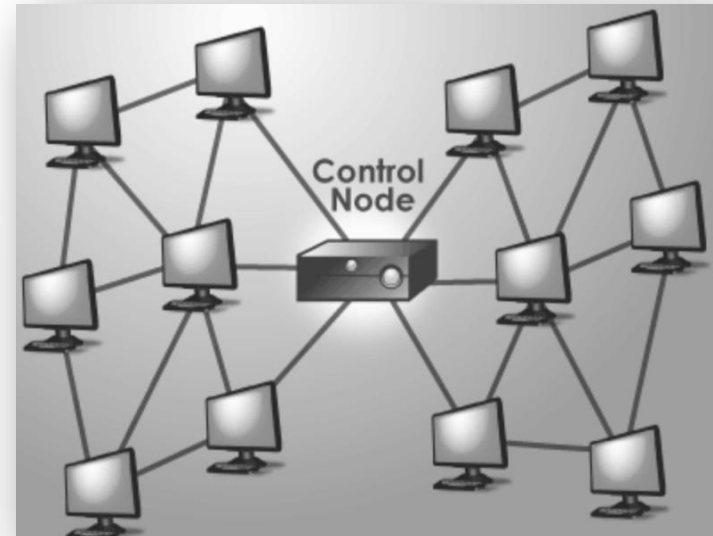
- **3 sous-groupes de SID:**
 - **Clusters de machines**
 - **Grilles de calcul**
 - **Clouds**

Catégories de SD

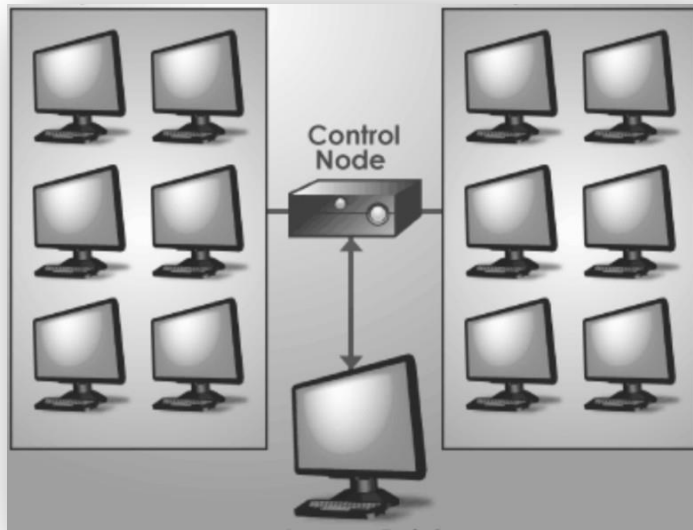
Systèmes informatiques distribués



Grille

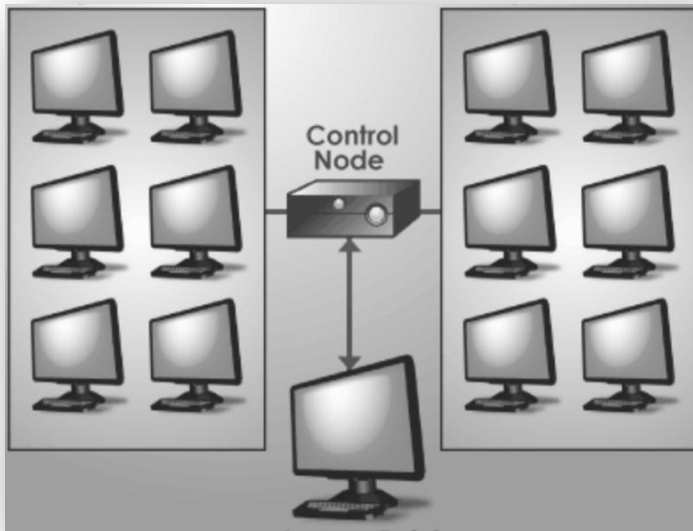


Cluster

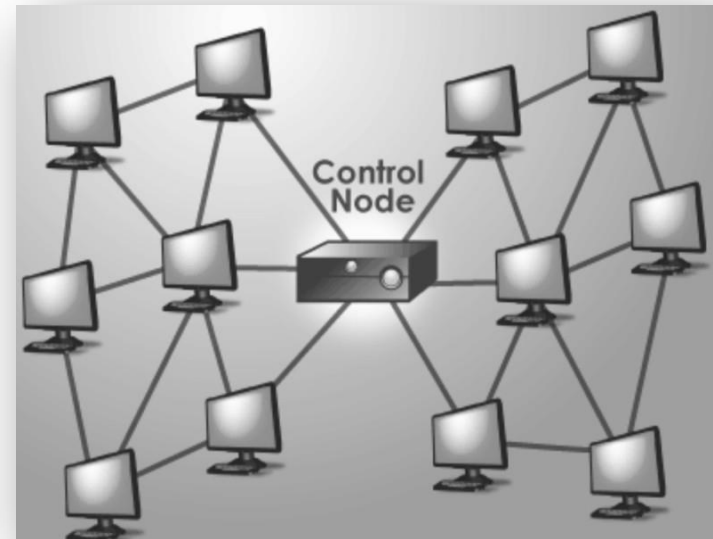


Catégories de SD

Systèmes informatiques distribués



- **Centralisé**
- **Plusieurs postes similaires (homogène)**
- **Même emplacement physique**
- **Liaison réseau haut-débit**



- **Distribué**
- **Plusieurs postes différents (hétérogène)**
- **Différents emplacements physiques**

Catégories de SD

Systèmes informatiques distribués



Récente catégorie



Catégories de SD

Systèmes informatiques distribués



Exercice d'application

Faites un tableau comparatif des **clusters**, **grilles** et **cloud**

Catégories de SD

Systèmes d'information distribués



- **Distribution de l'information sur différents composants d'un système distribué**
- **Exemple:**
 - **Base de données distribuées**
 - **Architecture 2, 3, n-tiers**
 - **Peer-to-peer vs Architecture centralisée client-serveur**

Systèmes d'informations distribués

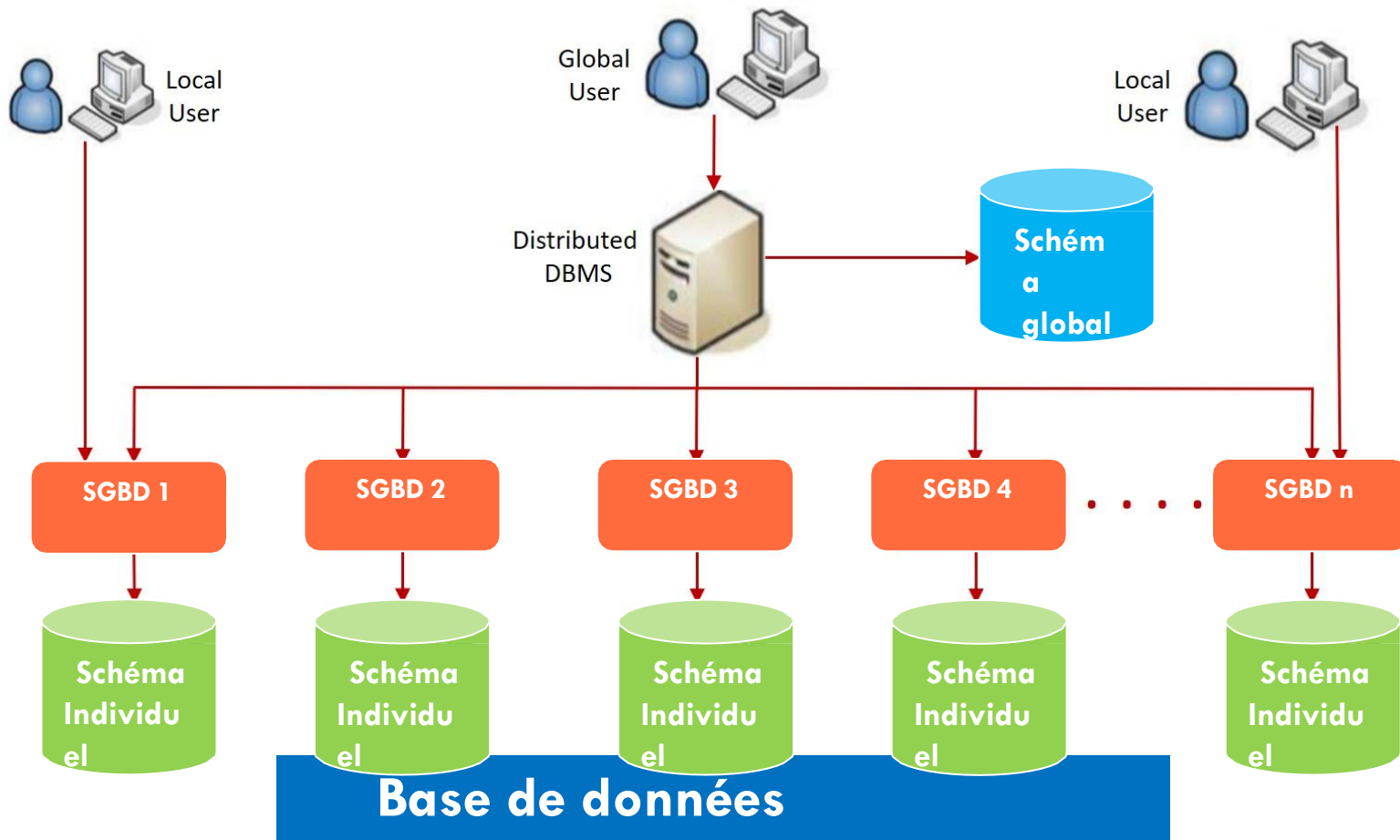
Base de données distribuées



- **Différents types de bases de données déployés à travers plusieurs zones géographiques**
- **Fédération de différents SGBDs**
- **Problème:**
 - **Comment définir un schéma de données global, de façon à garantir l'accessibilité, la transparence, et l'intégrité aux utilisateurs ?**

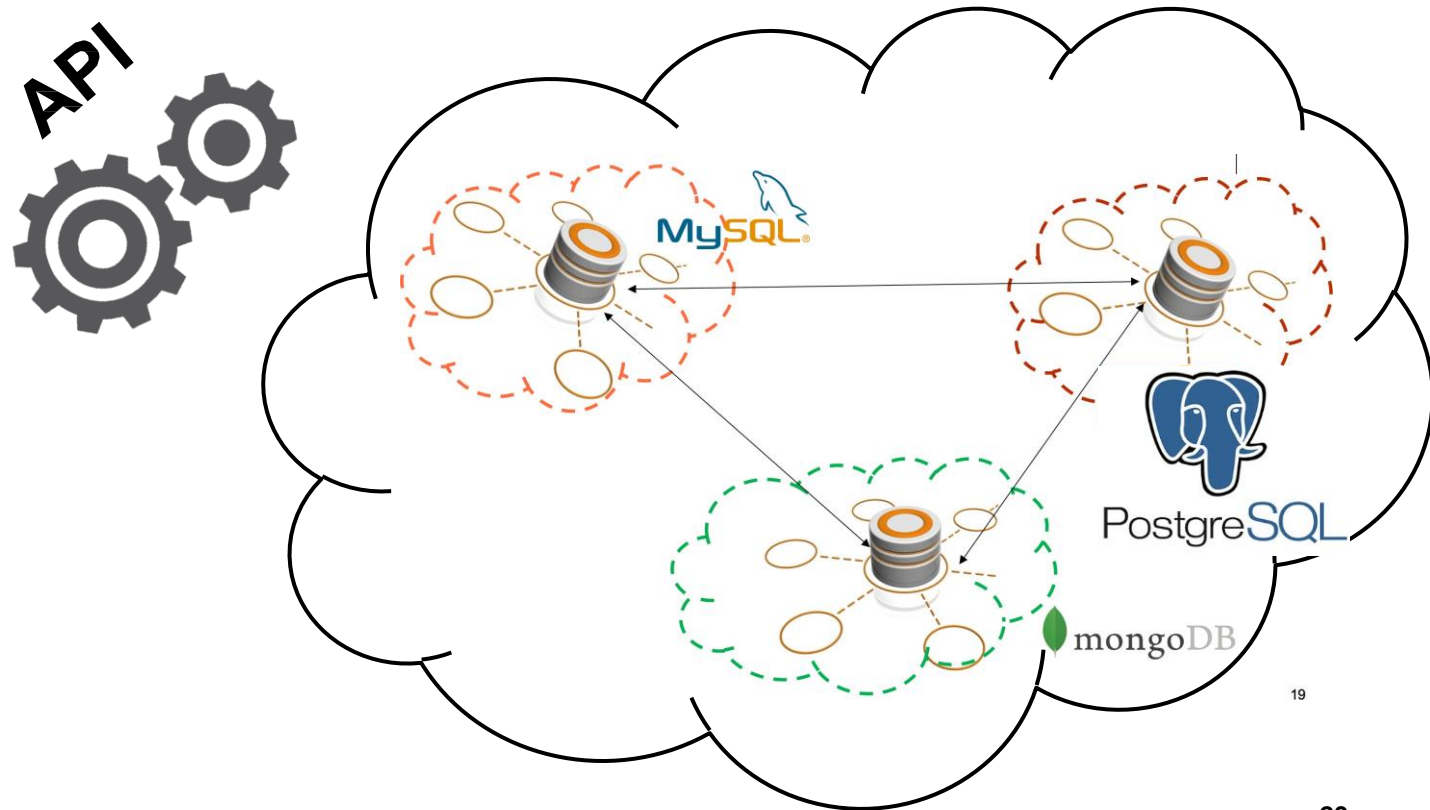
Systèmes d'informations distribués

Base de données distribuées



Systemes d'informations distribués

Base de données distribuées



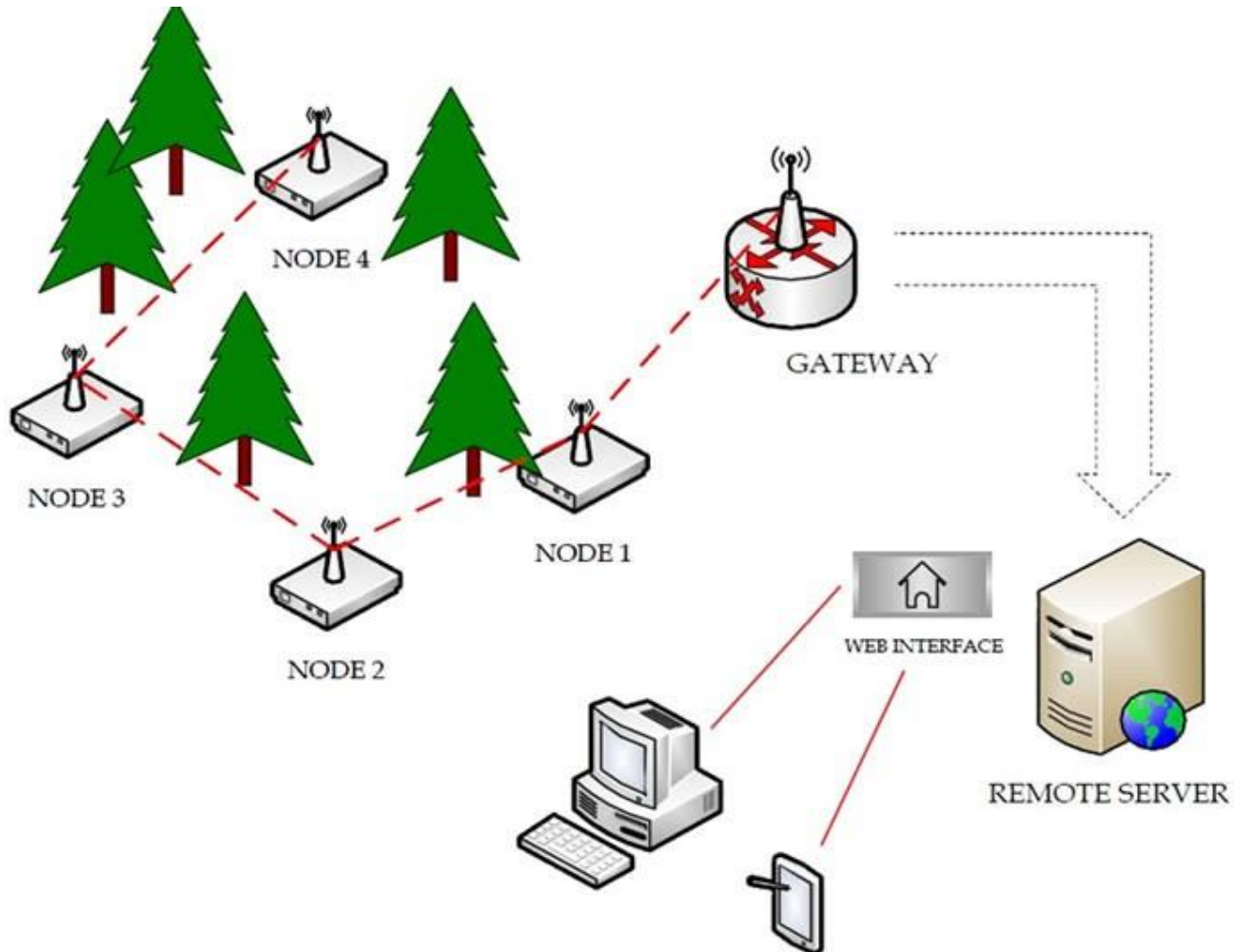
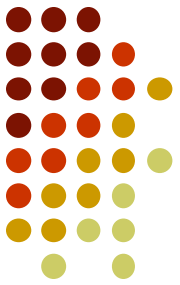
Catégories de SD

Systèmes embarqués et diffuses



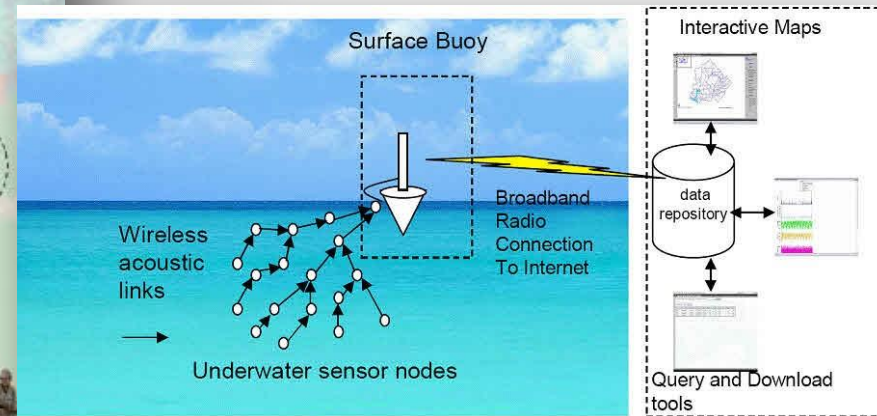
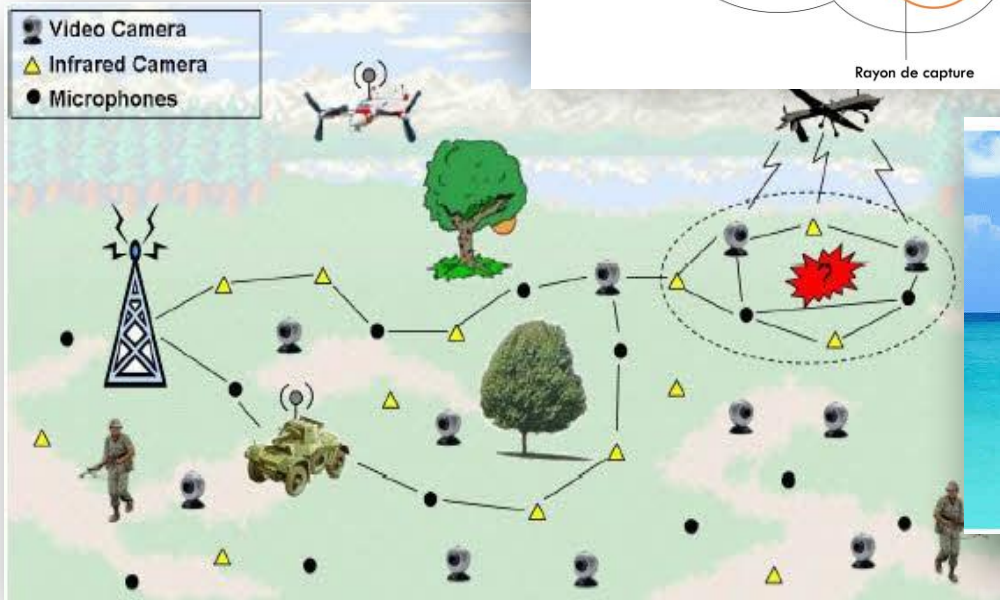
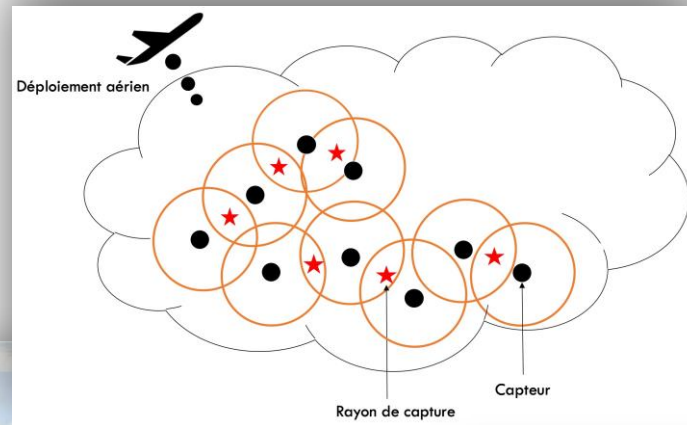
- ✓ **Système composé d'équipements électronique autonomes et aux ressources limitées, spécialisé dans une tâche bien précise.**
- ✓ **Les équipements sont souvent limités en taille (minuscules équipements équipés de microcontrôleur) et en énergie (consommation restreinte).**

Réseau de capteurs sans fil



Réseau de capteurs sans fil

Déploiement



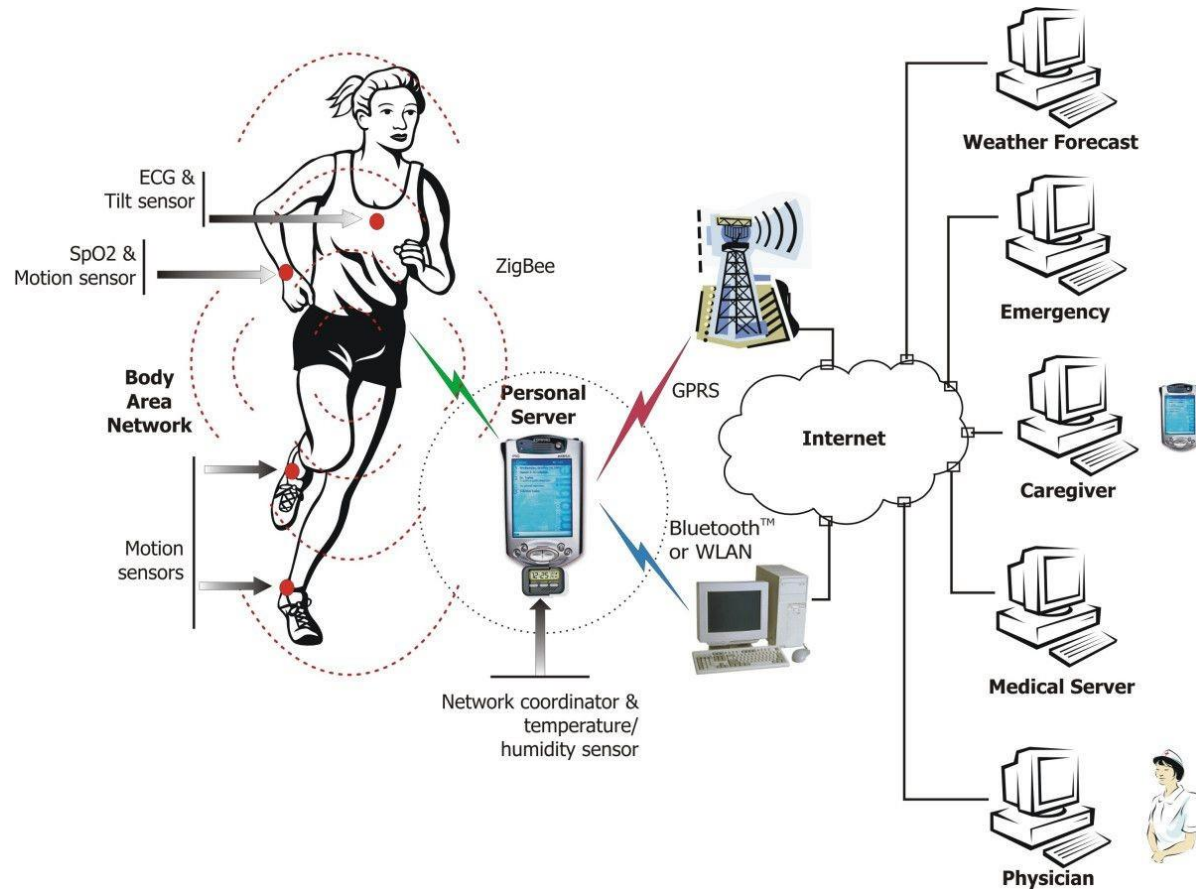
Réseau de capteurs sans fil

Home systems



Réseau de capteurs sans fil

Body area network (BAN)



Source: A wireless body area network of intelligent motion sensors for computer assisted physical rehabilitation
Emil Jovanov, Aleksandar Milenkovic, Chris Otto and Piet C de Groen. *Journal of NeuroEngineering and Rehabilitation* 2005,
2:6 <https://doi.org/10.1186/1743-0003-2-6> Jovanov et al; licensee BioMed Central Ltd. 2005

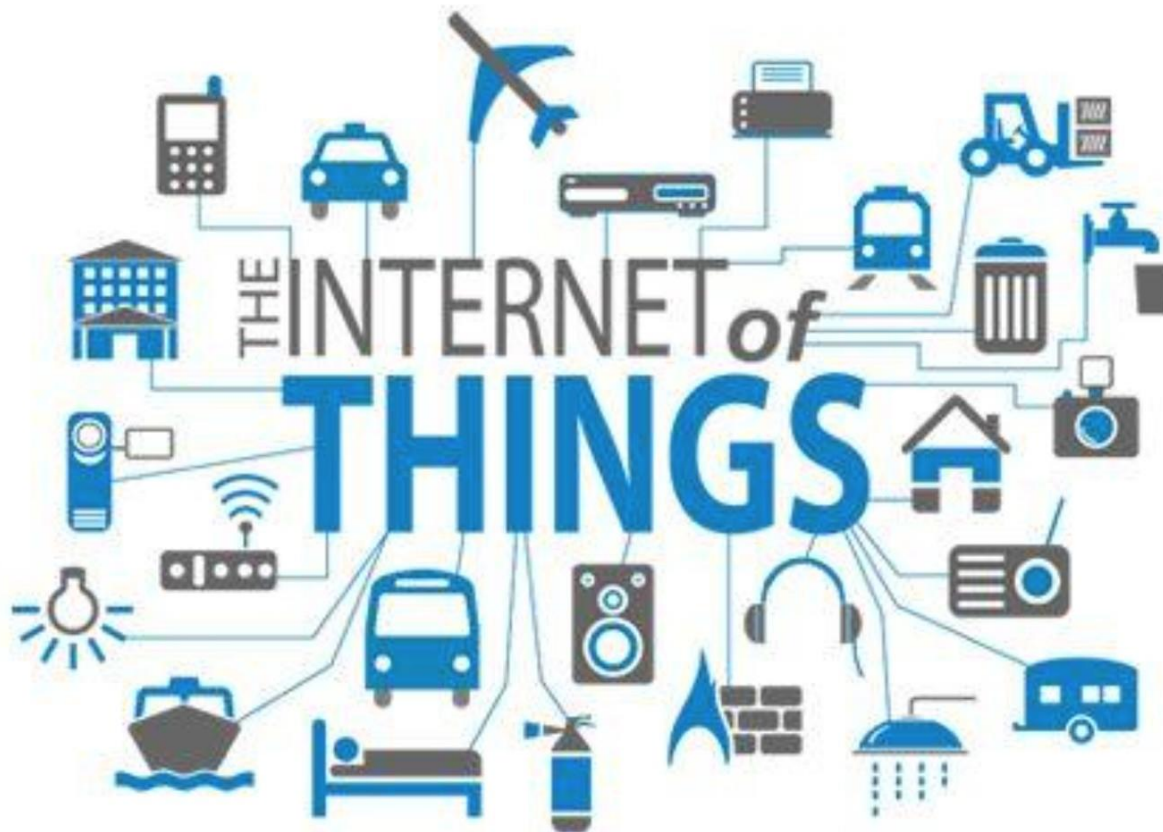
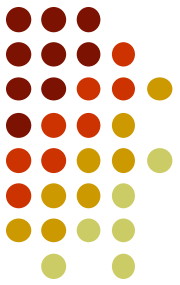
Réseau de capteurs sans fil

Smart city



Source https://www.proximus.be/fr/id_b_cl_ontwikkeling_van_smart_city/entreprises-et-secteur-public/decouvrir/news-blog/actualite/demarche-smart-city.html

Internet des objets



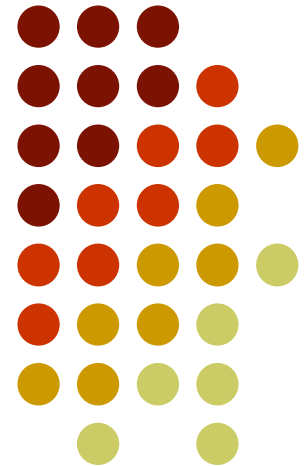
Kuva 1. Internet of Things. Lähde: Huffington Post

Internet des objets

Architecture



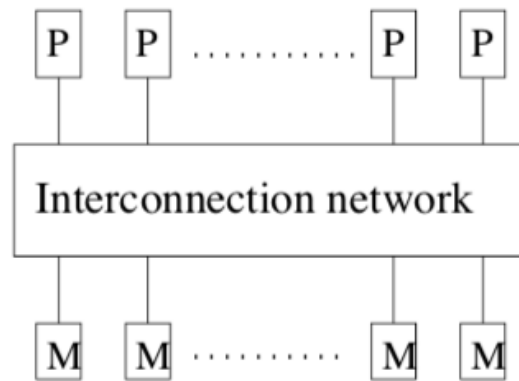
Système multiprocesseurs



Système multiprocesseurs



Un système multiprocesseur est un système parallèle composé de **plusieurs processeurs** ayant un accès direct à la **mémoire partagée** qui forme un espace d'adressage commun.



(a)

M memory

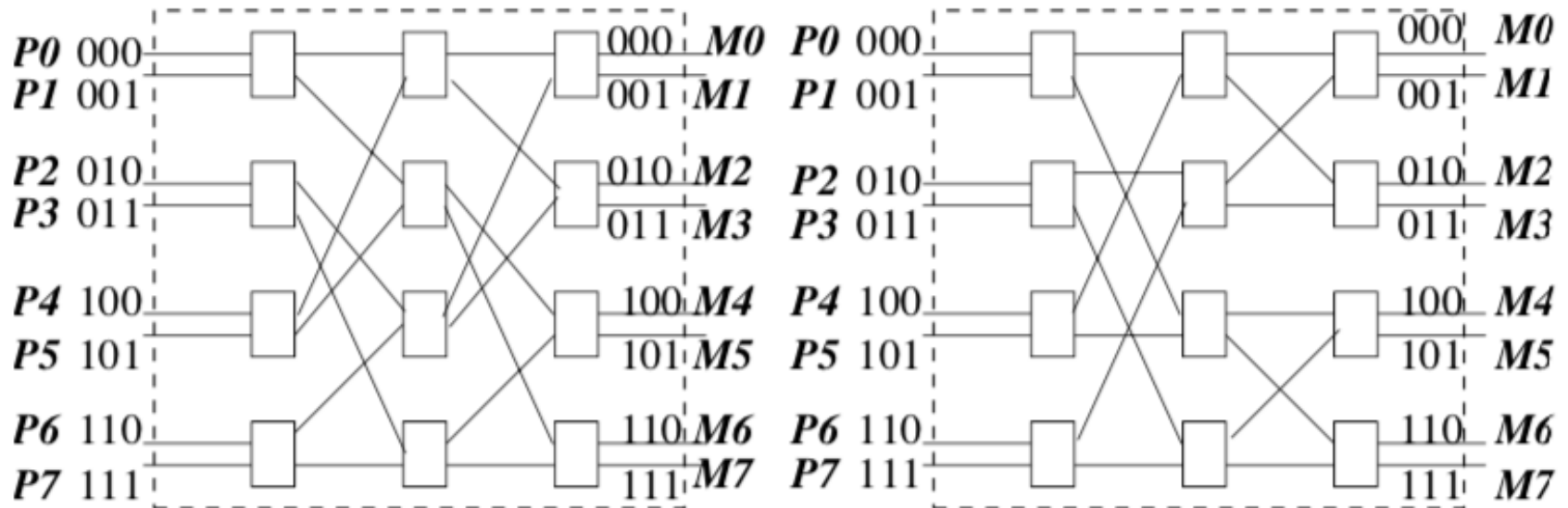
P processor

Système multiprocesseurs



- L'accès mémoire dans un système multiprocesseur est uniforme (UMA)
- Le temps de latence ou temps d'attente pour accéder à n'importe quel emplacement mémoire à partir de n'importe quel processeur du système est le même
- La communication interprocessus entre processeurs se fait par des opérations de lecture et d'écriture sur la mémoire partagée, ou par les routines MPI
- Tous les processeurs exécutent généralement le même système d'exploitation, et le matériel et le logiciel sont très étroitement liés.

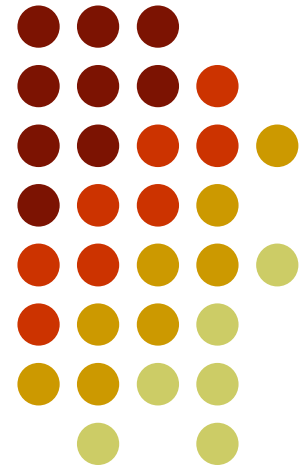
Interconnexion de processeurs



(a) 3-stage Omega network ($n=8, M=4$)

(b) 3-stage Butterfly network ($n=8, M=4$)

Systeme parallele multi-ordinateurs

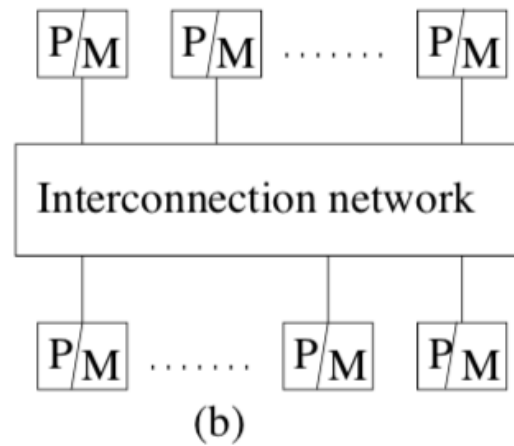


Système parallèle multi-ordinateurs



- Un système parallèle multi-ordinateurs est un système parallèle dont les multiples processeurs n'ont pas directement accès à la mémoire partagée.

**Appelé aussi
« Cluster d'ordinateurs »**



M memory

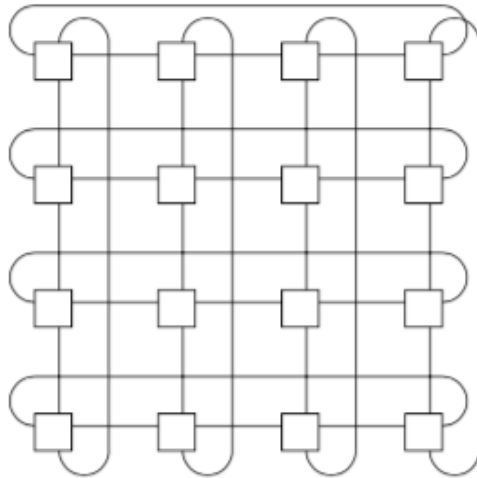
P processor

Systeme parallele multi-ordinateurs

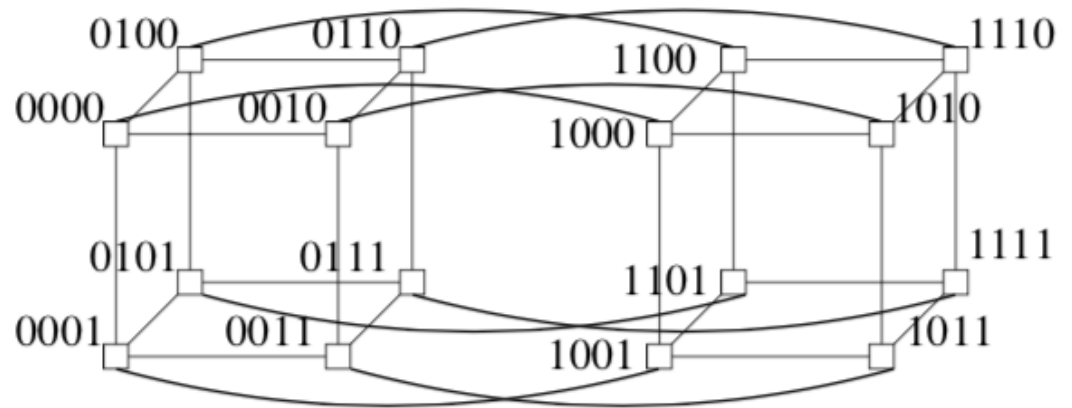


- La mémoire des multiples processeurs peut ou non former un espace d'adressage commun.
- Ces ordinateurs n'ont généralement pas d'horloge commune.

Interconnexion de machines



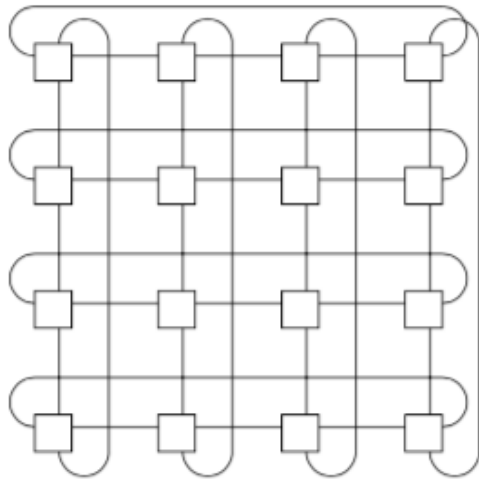
2D mesh array



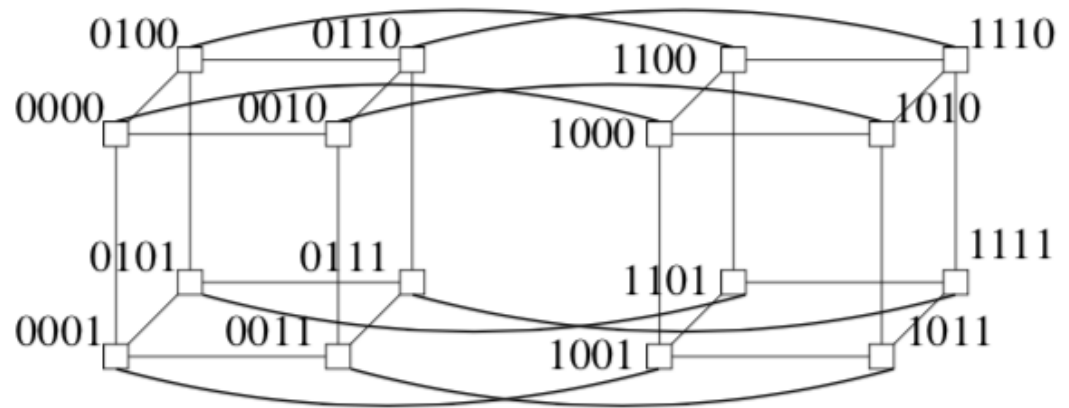
□ processor + memory

Hypercube

Interconnexion de machines



2D mesh array

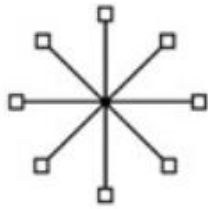


□ processor + memory

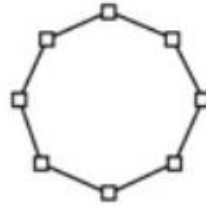
Hypercube

Important: Pour des processeurs reliés comme tels, retrouver un chemin à partir de n'importe quel couple de nœuds, est crucial. L'organisation a une incidence sur le routage, et l'accessibilité des données

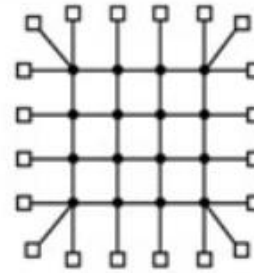
Interconnexion de machines



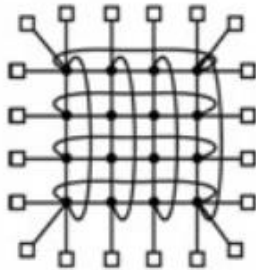
Etoile



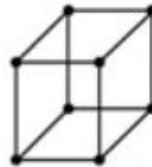
Anneau



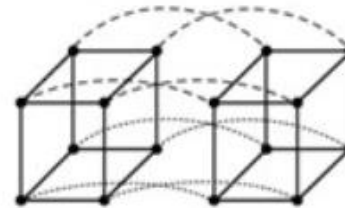
Grille



2D mesh array



Cube



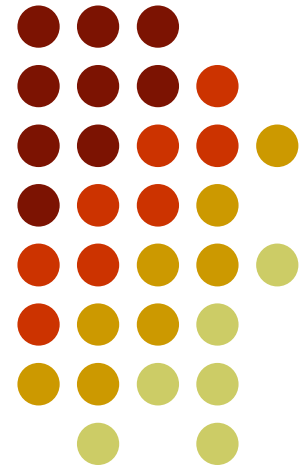
hypercube



SD vs Système multiprocesseurs vs Système multi-ordinateurs

Item	Multiprocesseur	Multi-ordinateurs	SD
Configuration nœud	CPU	CPU, RAM, Réseau	Ordinateur
Périphériques	Totalement partagé	Partagé ou non	Mémoire locale
Emplacement	Même circuit	Même pièce	Séparé géographiquement
Communication entre nœuds	RAM partagé	Interconnexion dédiée	Réseau traditionnel
OS	Un seul/partagé	Plusieurs/même	Tous différents
Système de fichiers	Un seul/partagé	Un seul/partagé	Propre à chaque nœud
Administration	1 organisation	1 organisation	1 organisation

Taxinomie de Flynn





Taxinomie de Flynn

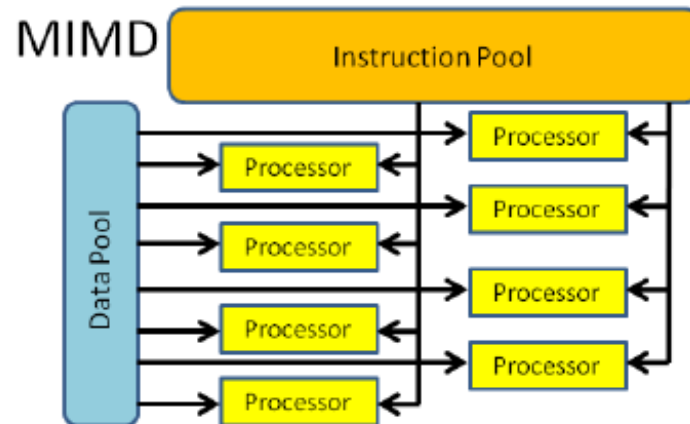
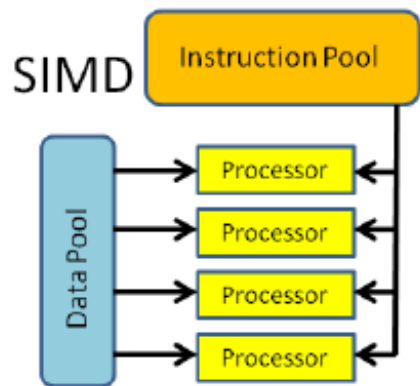
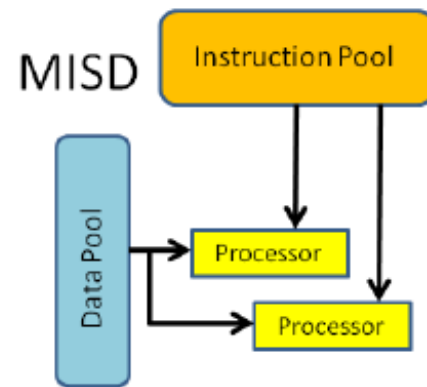
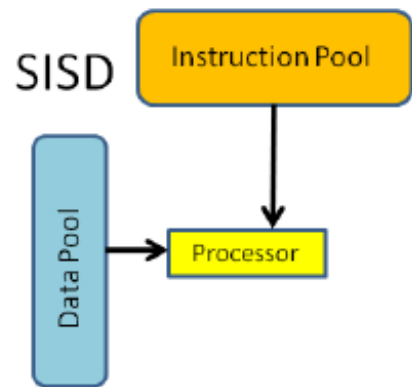
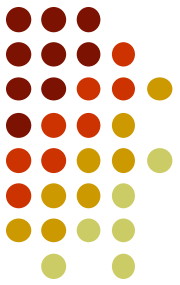
- Beaucoup de tentatives ont été élaborées pour catégoriser les types d'ordinateurs
- Flynn est allé bien loin en définissant 4 catégories par rapport au flot d'instructions et aux flots de données



Taxinomie de Flynn

- SISD: Single Instruction Single Data -Von Neumann
- SIMD: Single Instruction Multiple Data
- MISD: Multiple Instruction Single Data
- MIMD: Multiple Instruction Multiple Data

Taxinomie de Flynn



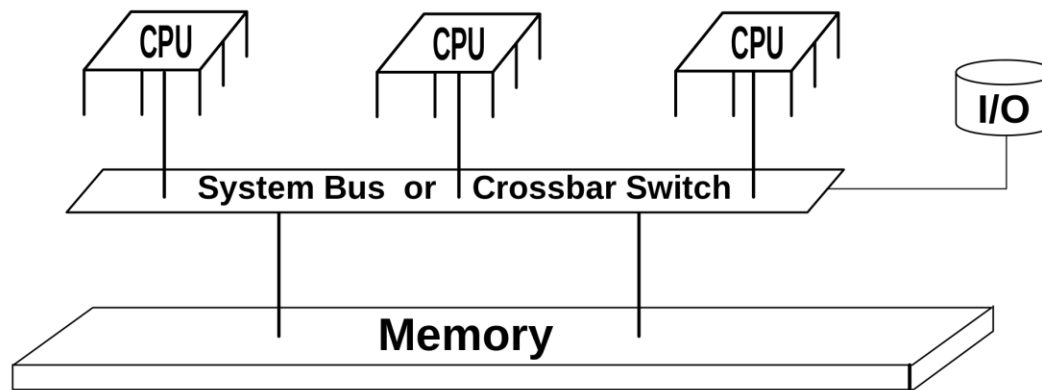
Problèmes derrière la Taxinomie de Flynn



Problèmes

- Inutilité de MISD
- Non homogénéité du parallélisme
- MIMD inclut beaucoup de sous-catégories
 - Mémoire partagé ou distribuée
 - Bus ou Switch

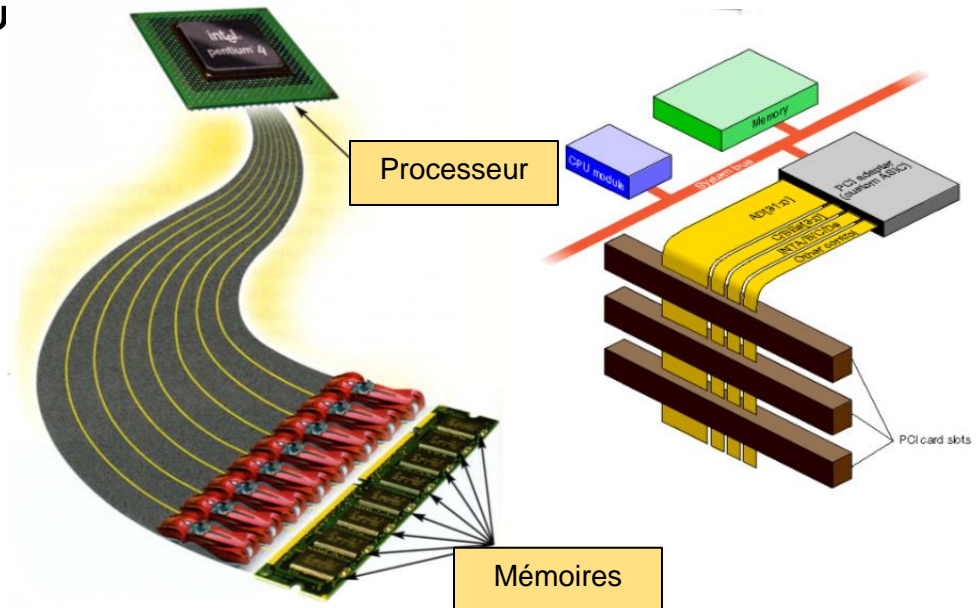
Mémoire partagée



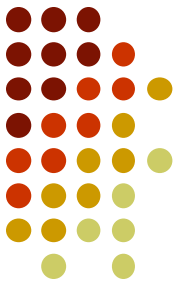
La mémoire partagée est une mémoire à laquelle peuvent accéder simultanément plusieurs programmes dans le but de **communiquer entre eux** ou d'**éviter les copies redondantes**.

Bus (1)

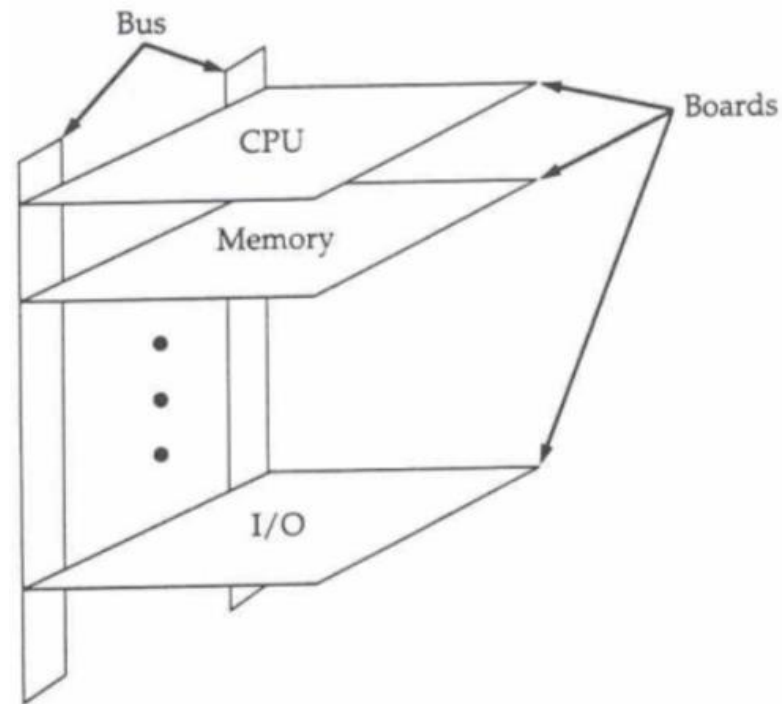
- **Un bus** est une voie de communication reliant deux ou plusieurs appareils.
- Une caractéristique clé d'un bus est qu'il s'agit d'un support de transmission partagé.
- **Un bus** se compose de plusieurs voies ou lignes. Chaque ligne est capable de transmettre un signal représentant un chiffre binaire (1 ou 0)



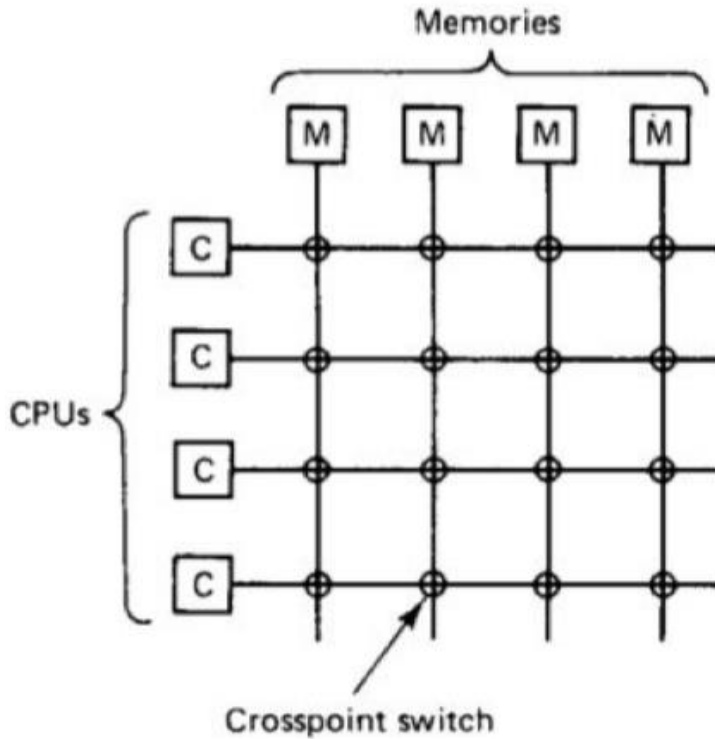
Bus (2)



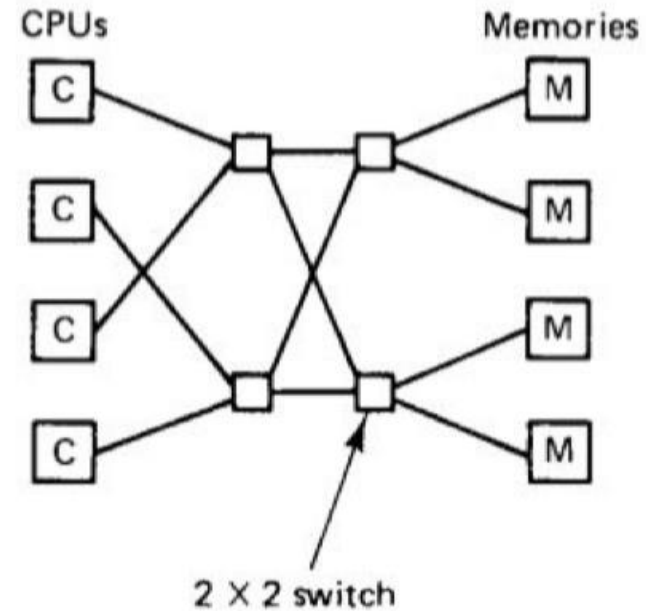
- **Le bus** système est un certain nombre de conducteurs électriques parallèles.
- Les conducteurs sont des lignes métalliques fixées sur une carte ou une carte de circuit imprimé.
- **Le bus** s'étend sur tous les composants qui se connectent aux lignes de bus.



Switch (1)



Grille de commutation



Réseau Omega

Switch (2)

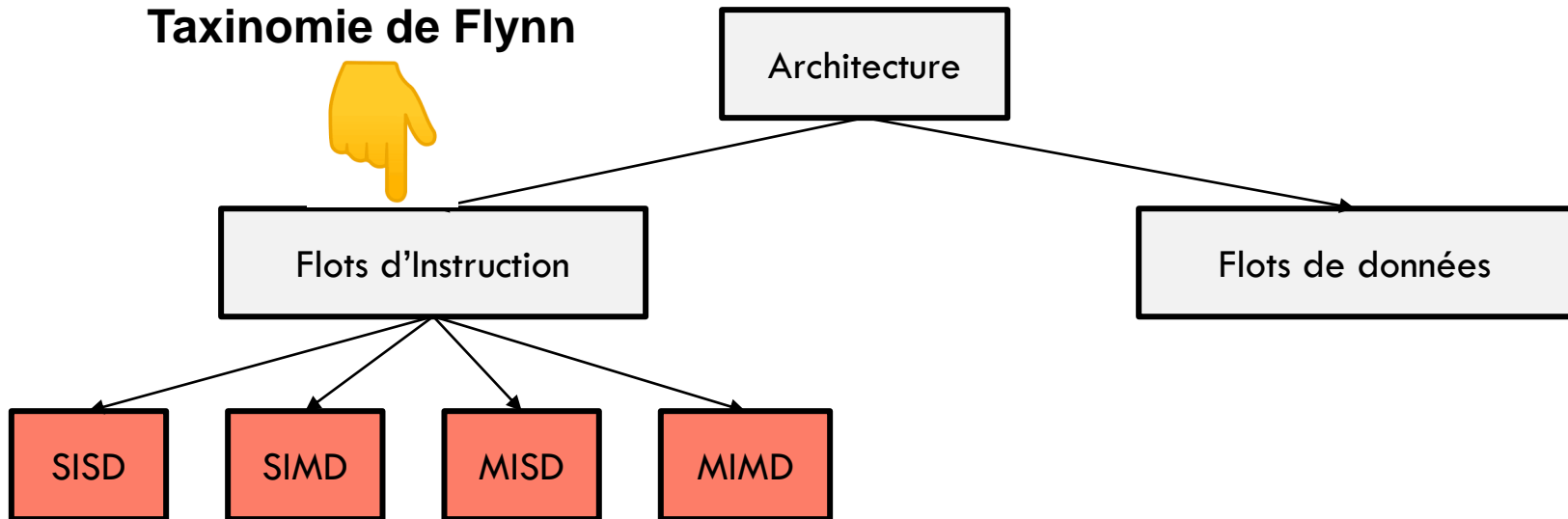


- À chaque intersection se trouve un minuscule commutateur de point de croisement électronique qui peut être ouvert et fermé dans le matériel.
- Lorsqu'un CPU souhaite accéder à une mémoire particulière, le commutateur de point de connexion les reliant est fermé pour permettre l'accès.
- Si deux CPU tentent d'accéder simultanément à la même mémoire, l'un d'eux devra attendre.

Extension de Flynn



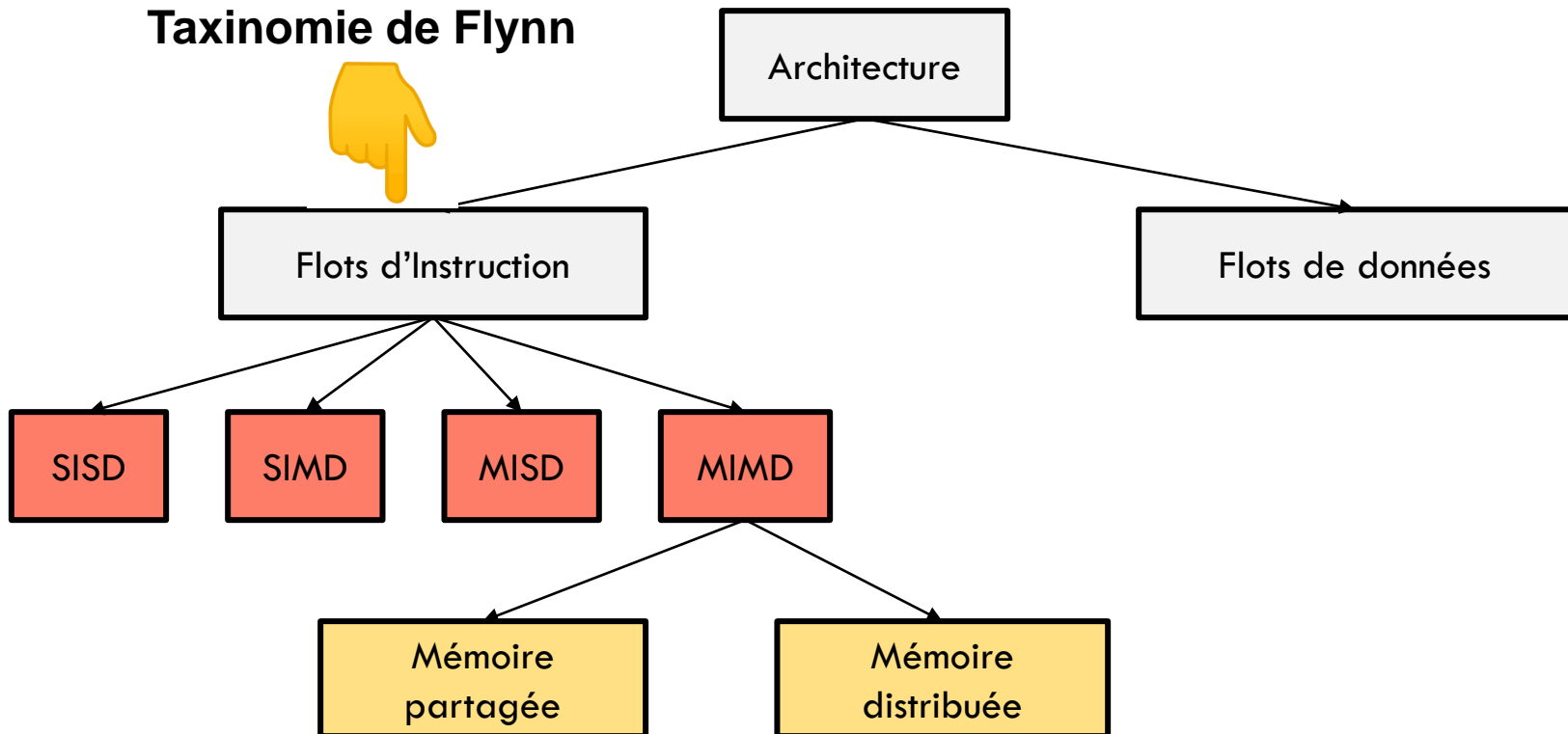
Taxinomie de Flynn



Extension de Flynn



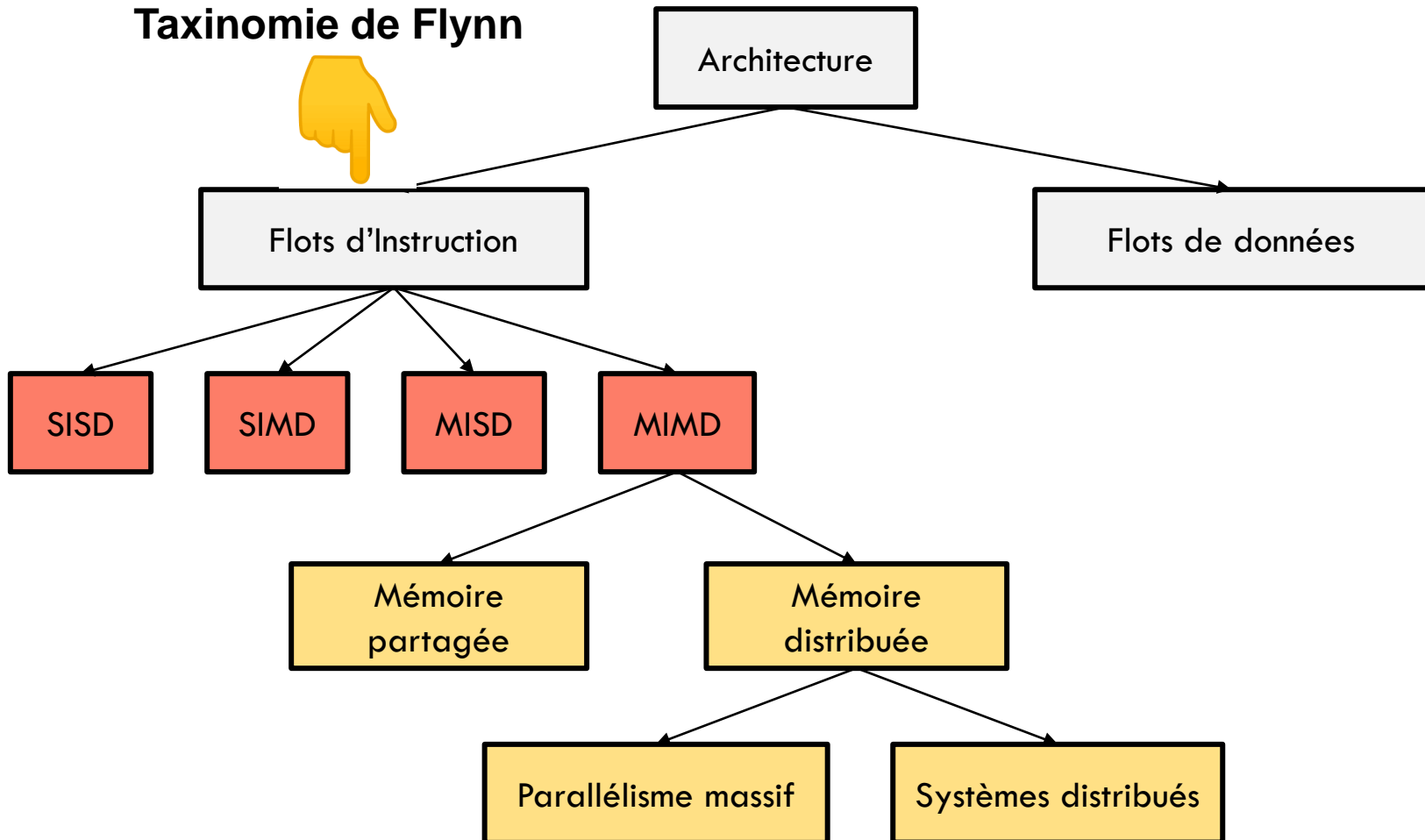
Taxinomie de Flynn



Extension de Flynn



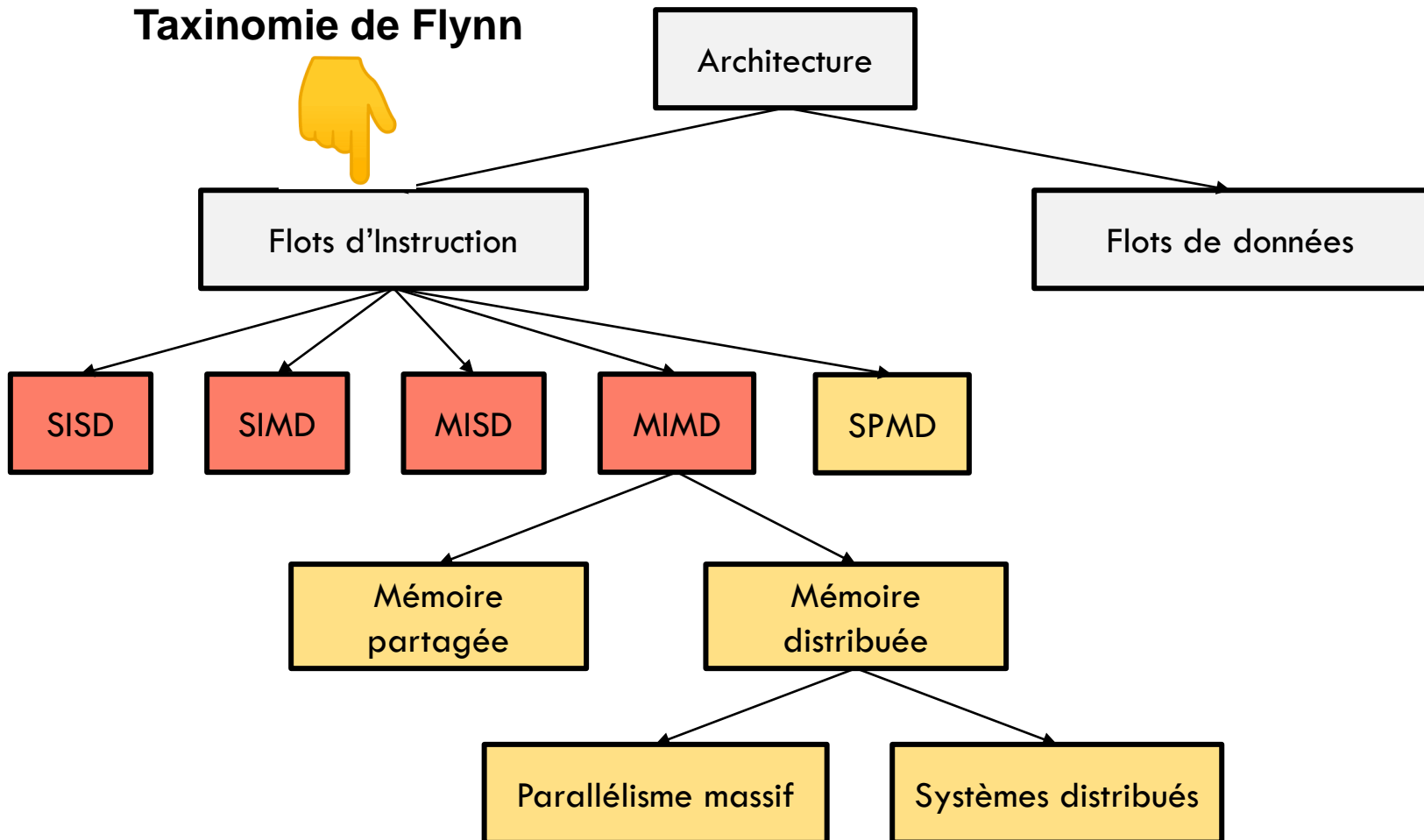
Taxinomie de Flynn



Extension de Flynn



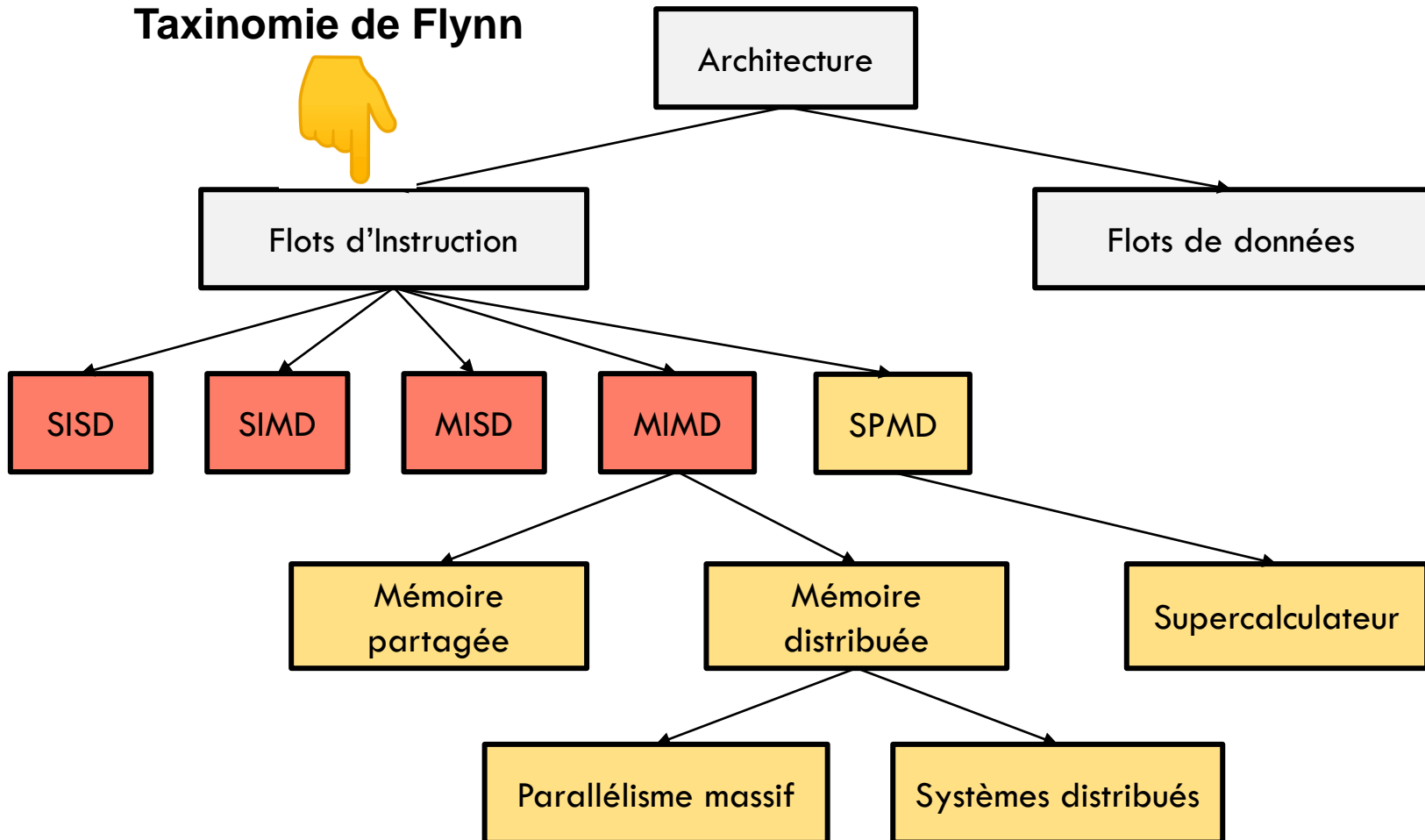
Taxinomie de Flynn



Extension de Flynn



Taxinomie de Flynn

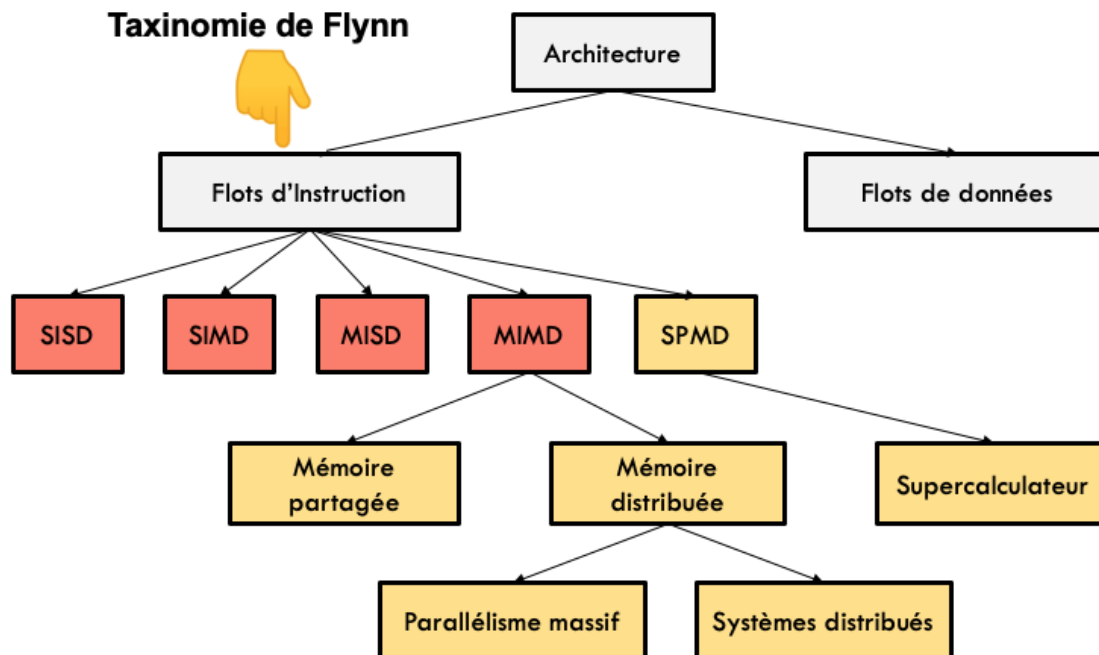




Extension de Flynn

SPMD : Single Program Multiple Data

- Exécuter différentes tâches d'un même programme
- Modèle de calcul des supercalculateurs



Conclusion



- **Les systèmes distribués sont constitués d'entités autonomes qui fonctionnent ensemble et donnant l'apparence d'un seul système cohérent**
- **Objectifs étant de garantir l'accessibilité, la transparence, l'ouverture et le passage à l'échelle**
- **Différentes catégories de SDs**
- **Parallélisme permet de mieux aborder les problèmes de calcul**