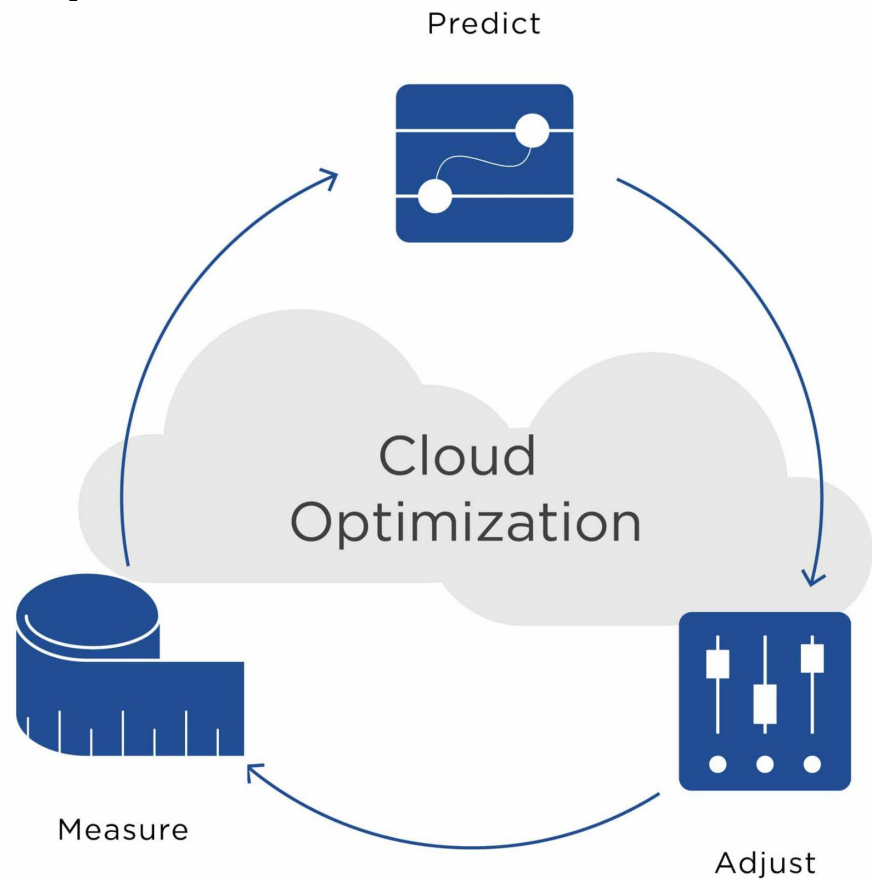


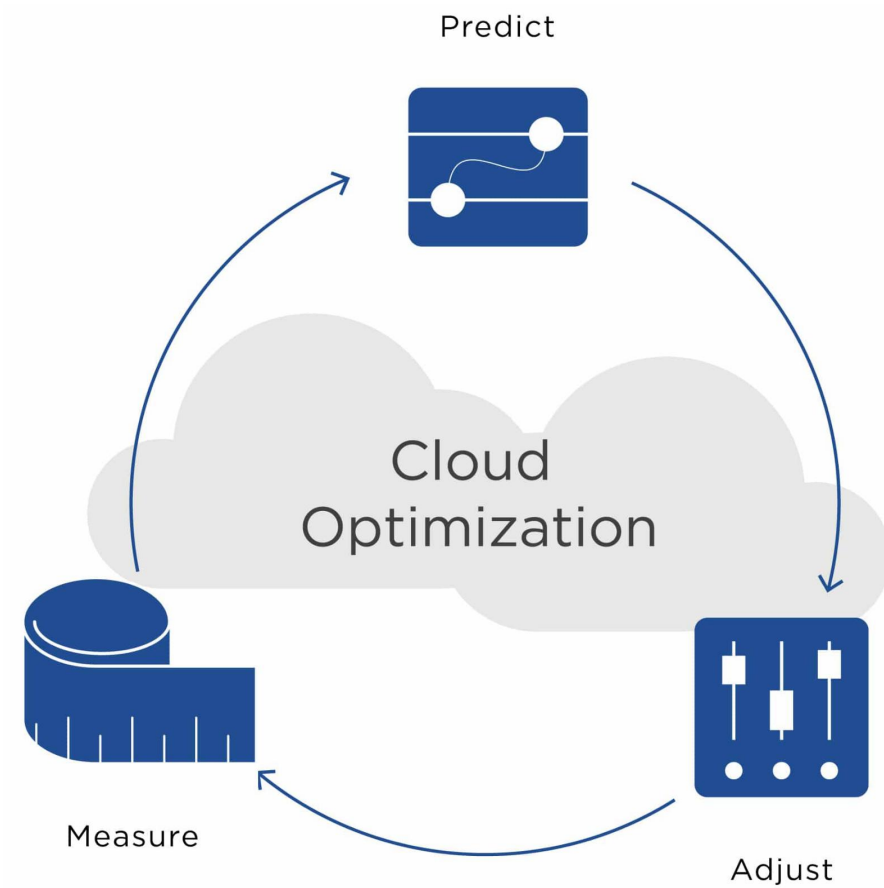
# PRIM

Optimisation d'une infrastructure de cloud par l'analyse prédictive des usages

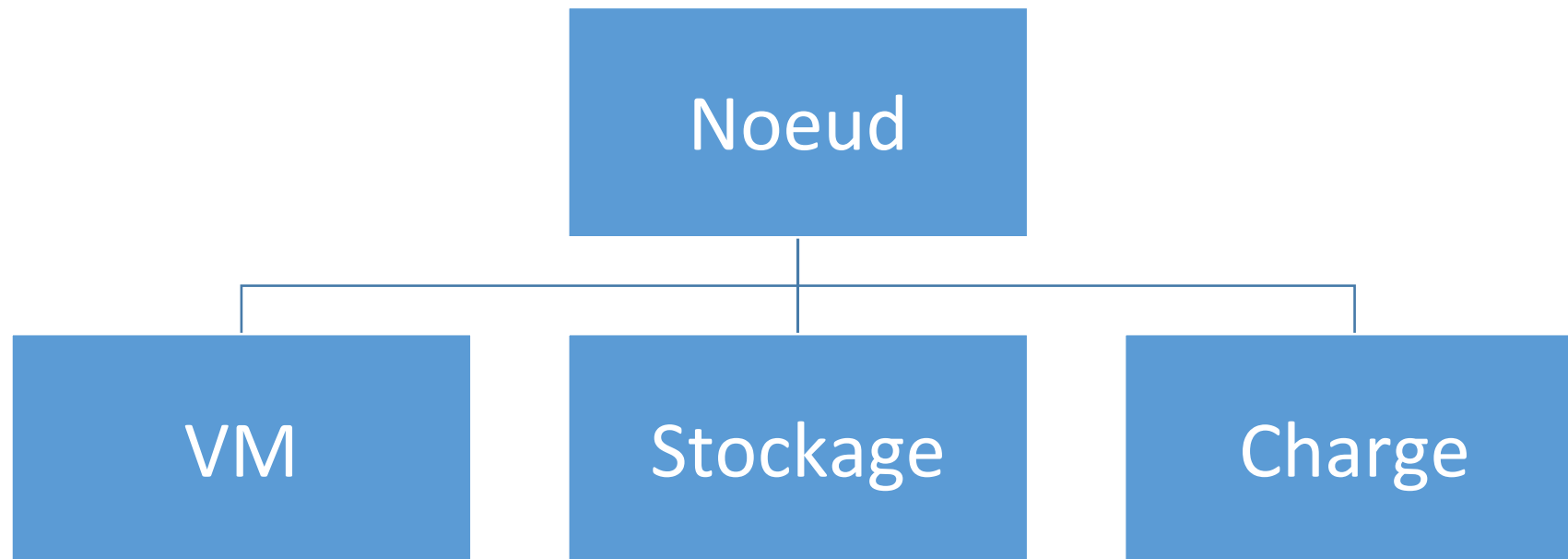


Mohamed Elhedi Ben Yedder

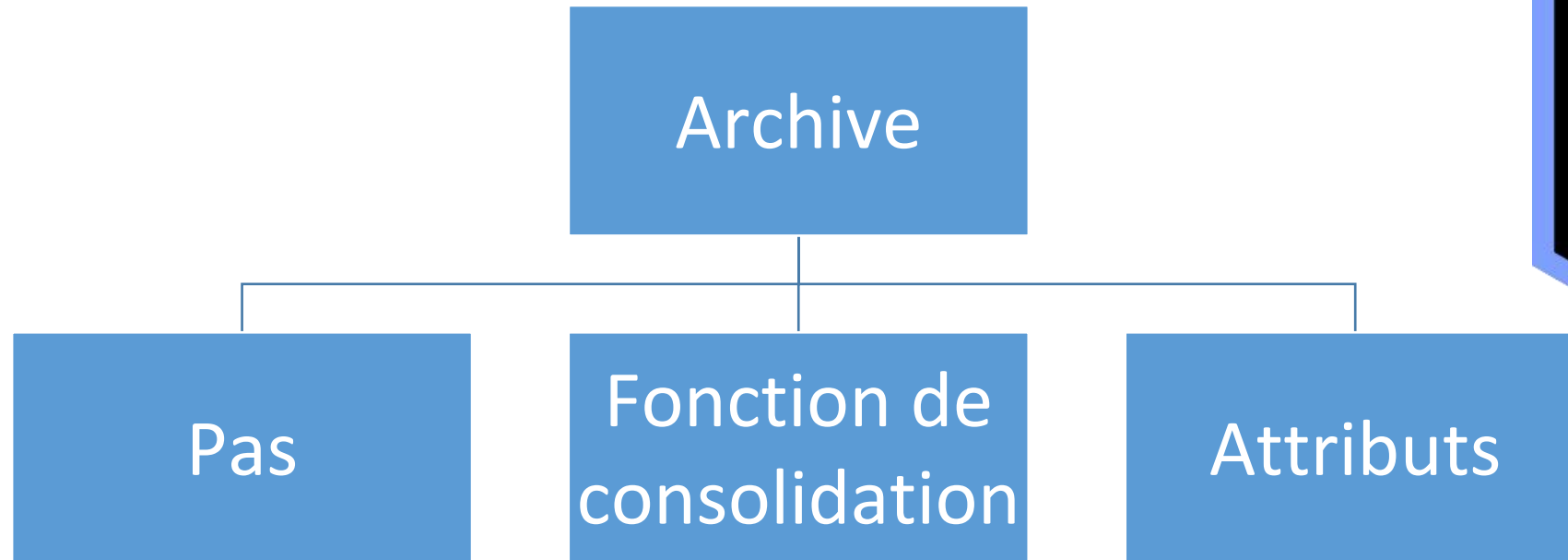
# Introduction



# *Description des données*

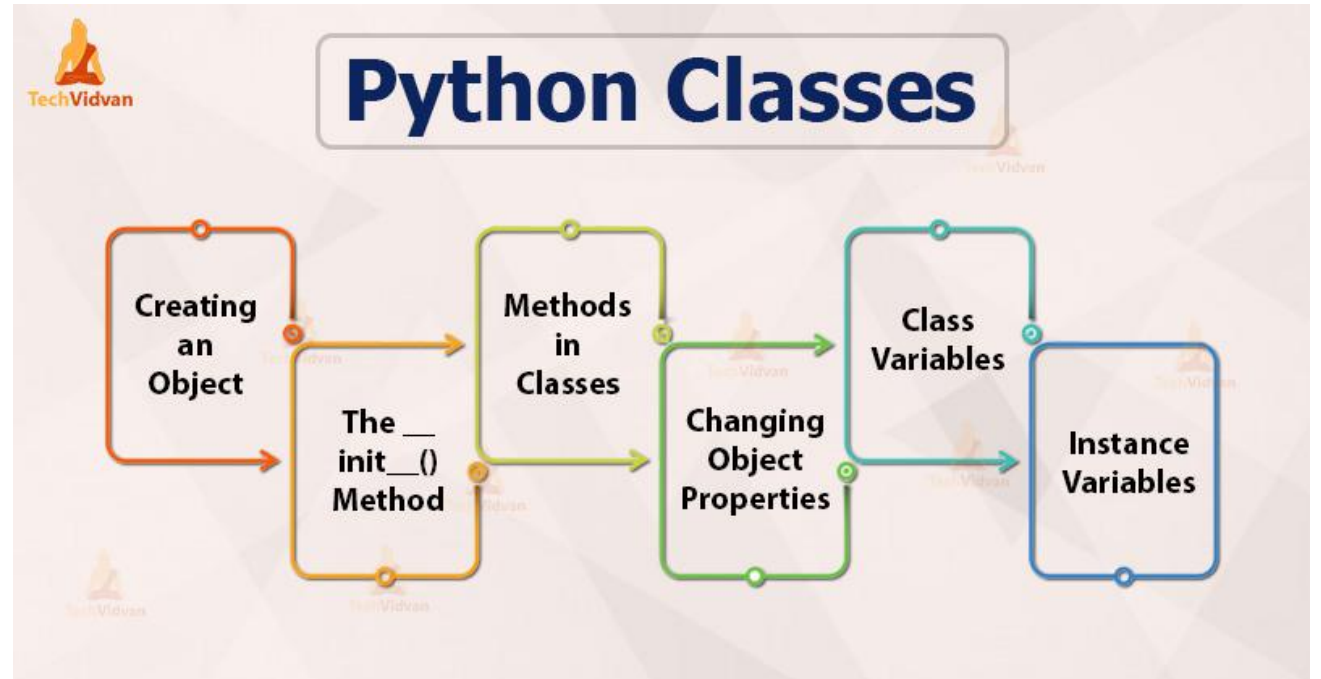


# *Description des données*

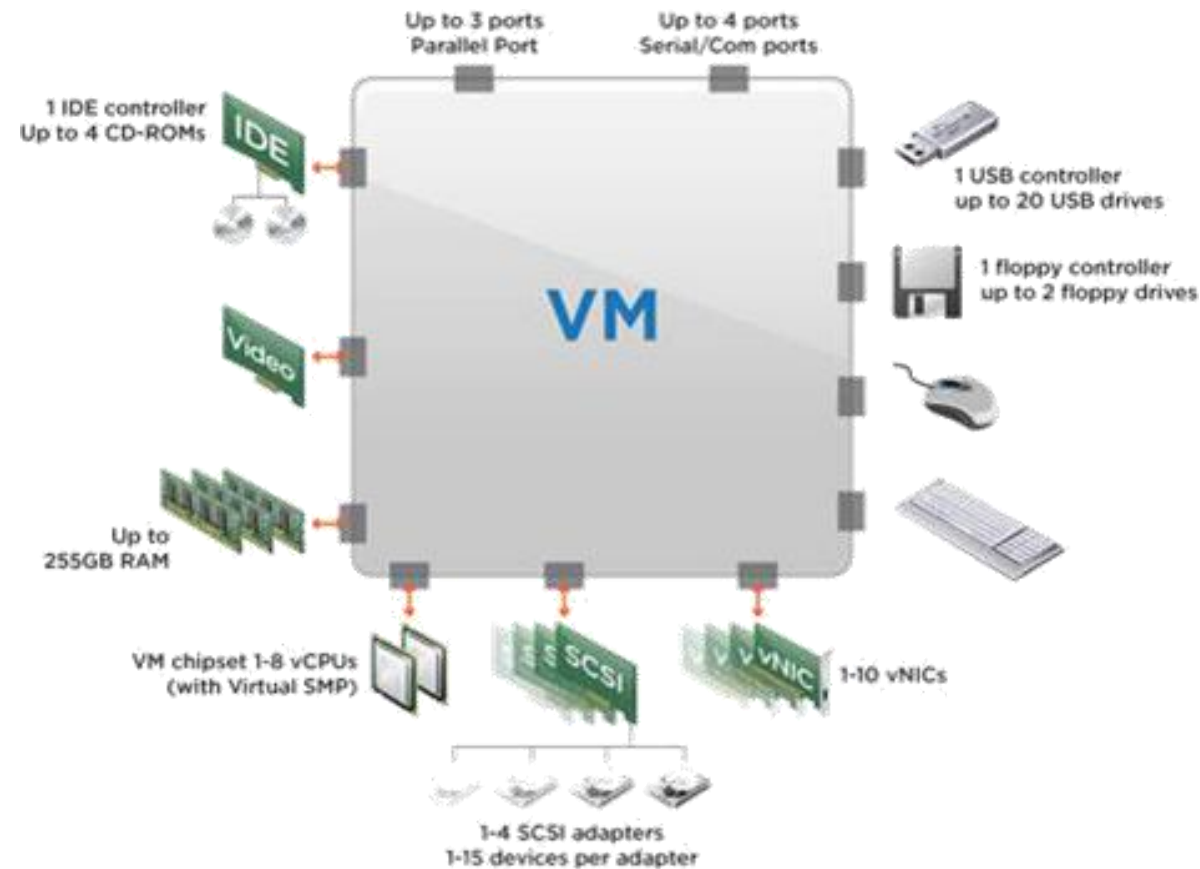


# Structuration de données

- RRD et XMLRRD
- VM
- Node
- Platform



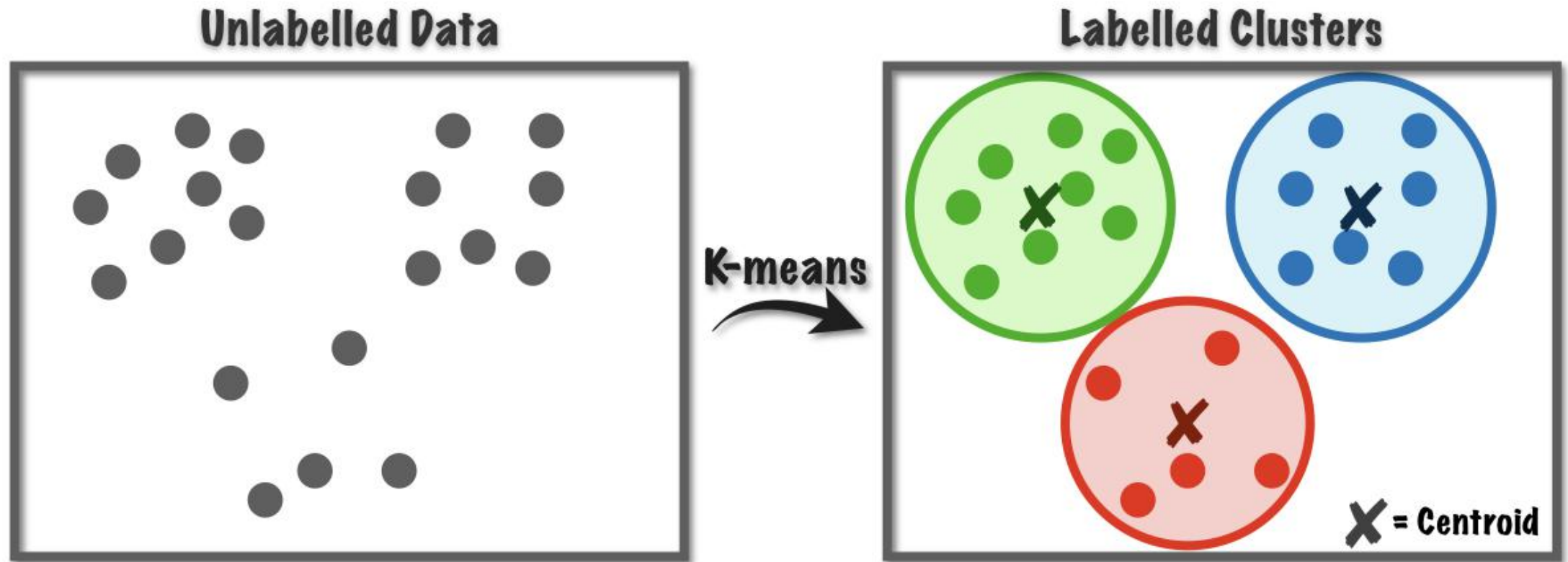
# Modélisation de la consommation d'une machine virtuelle



# Labellisation/Regroupement

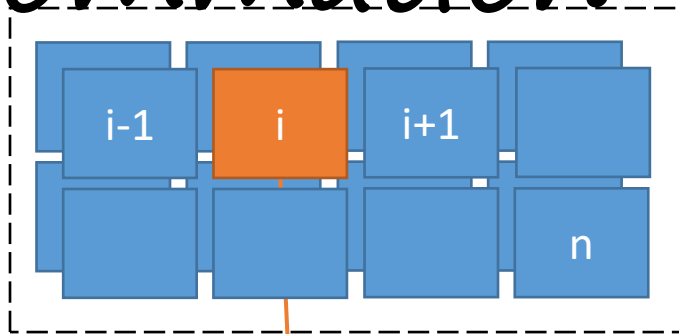
- $pas = 3heures$

- $k = 3$



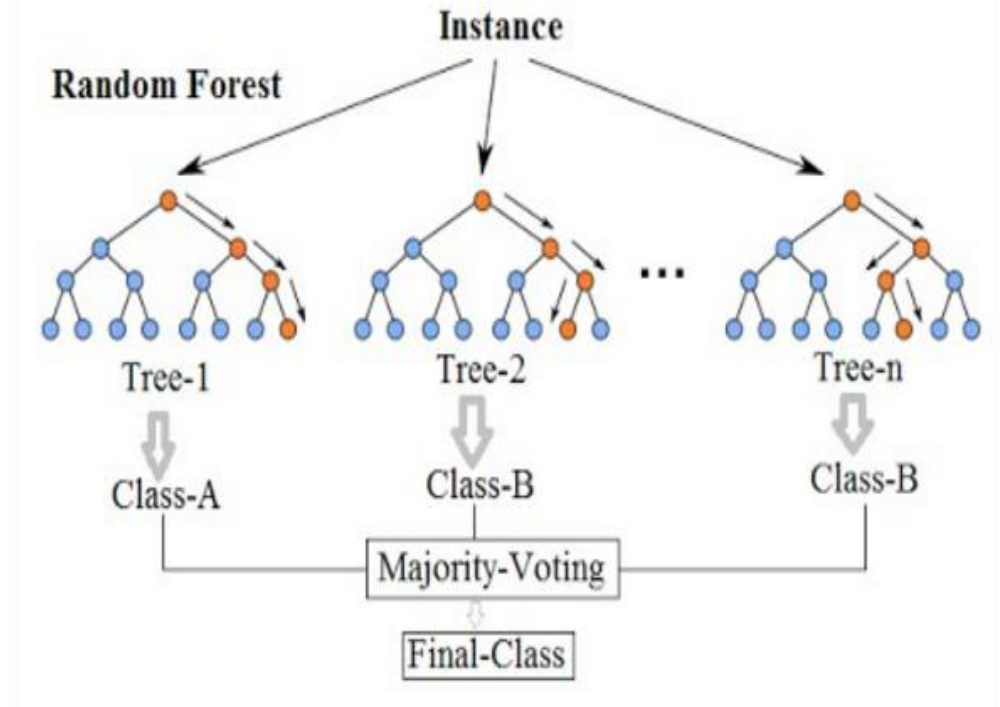
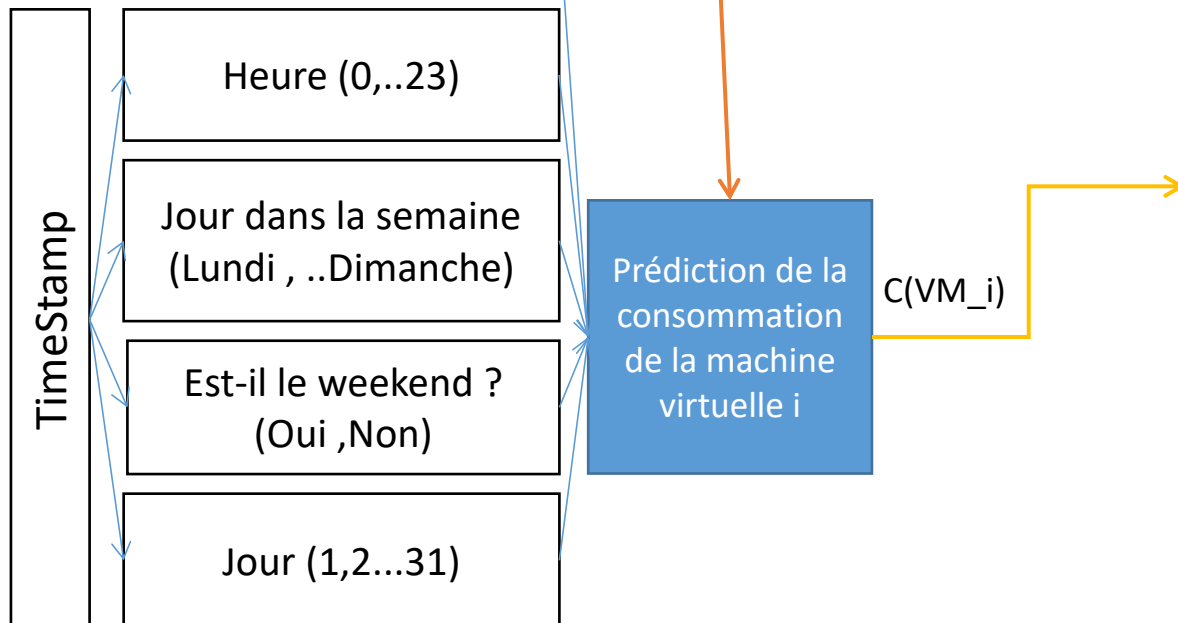
# Modèle pour prévoir la consommation

Les modèles des différentes machines virtuelles



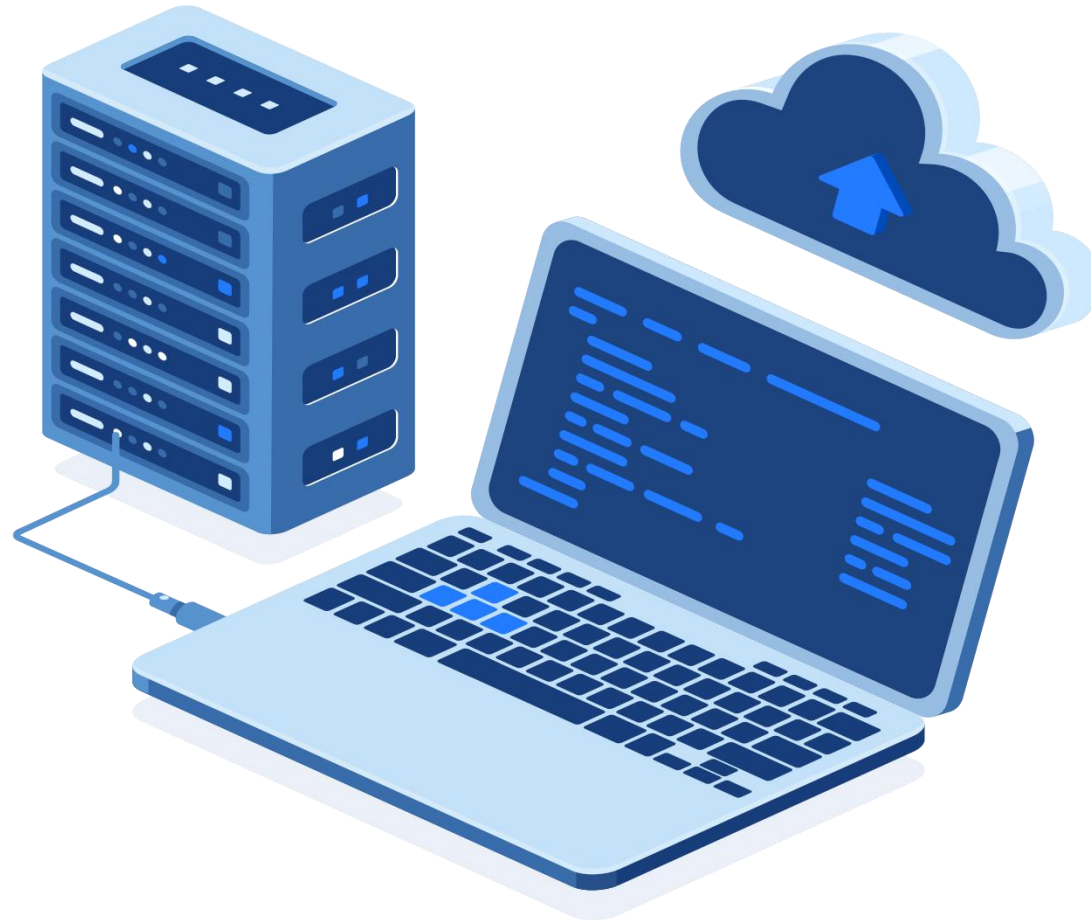
Nouvelle machine virtuelle à instancier de l'utilisateur  $i$

Les paramètres du modèle de la consommation de la machine virtuelle  $i$





# Modélisation de la charge d'un noeud



# Modélisation de la charge d'un noeud

- Données

$$I_p = \{vm_1, ..vm_p\}$$

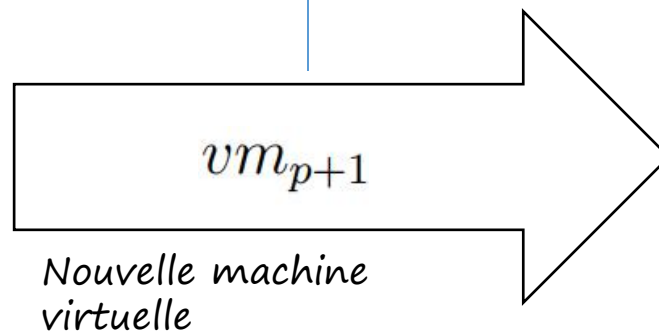
$$C(N, I_p, t_{p+1})$$

$$C(vm_{p+1}, t_{p+1})$$

- Objective

$$I_{p+1} = I_p U \{vm_{p+1}\}$$

$$C(N, I_{p+1}, t_{p+1}) = ?$$



# Modélisation de la charge d'un noeud

- Regression linéaire :

$$C(N, I_{p+1}, t_{p+1}) = C(N, I_p, t_{p+1}) + aC(vm_{p+1}, t_{p+1}) + b$$

- Changement de l'équation :

$$C(N, I_{p+1}, t) = \alpha \sum_{vm \in I_{p+1}} C(vm, t) + \beta(p+1) + \gamma$$

- Identification des coefficients :

$$a = \alpha, b = \beta$$

# Modélisation de la charge d'un noeud

- Détails de calcul :

$$C(N, I_{p+1}, t_{p+1}) = C(N, I_p, t_{p+1}) + aC(vm_{p+1}, t_{p+1}) + b$$

$$C(N, I_{p+1}, t) = \alpha \sum_{vm \in I_{p+1}} C(vm, t) + \beta(p+1) + \gamma$$

$$C(N, I_{p+1}, t) = \alpha(\sum_{vm \in I_p} C(vm, t) + C(vm_{p+1}, t)) + \beta(p+1) + \gamma$$

$$C(N, I_{p+1}, t) = (\alpha \sum_{vm \in I_p} C(vm, t) + \beta(p) + \gamma) + \alpha C(vm_{p+1}, t) + \beta$$

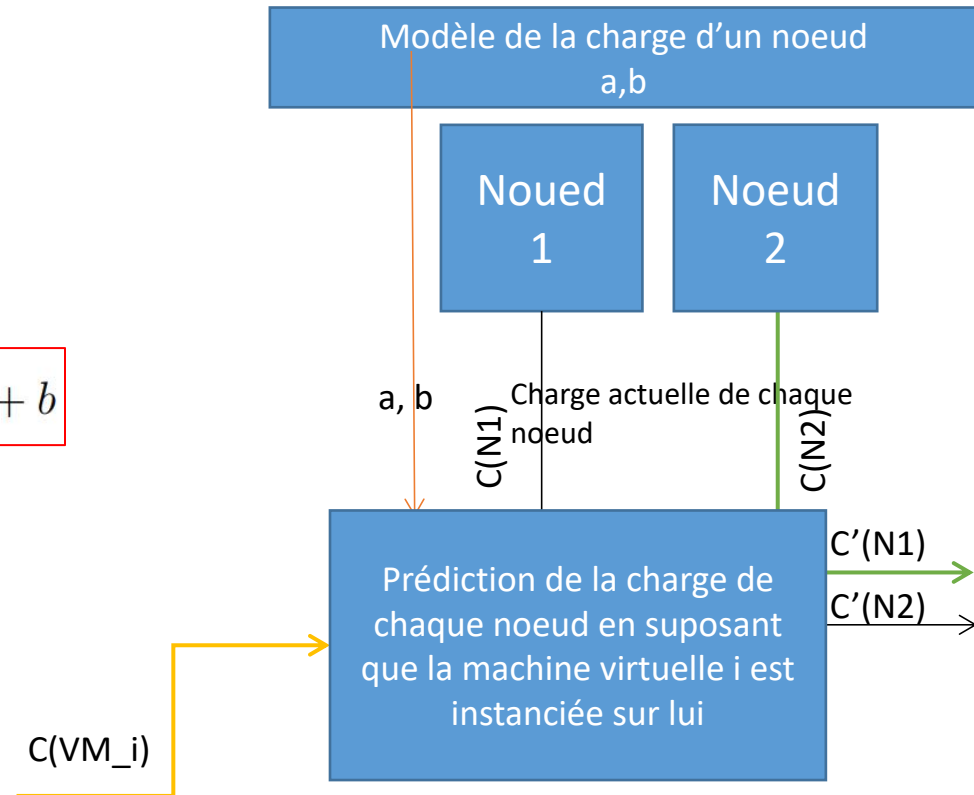
$$C(N, I_{p+1}, t) = C(N, I_p, t) + \alpha C(vm_{p+1}, t) + \beta,$$

$$C(N, I_{p+1}, t_{p+1}) = C(N, I_p, t_{p+1}) + \alpha C(vm_{p+1}, t_{p+1}) + \beta$$

# Modélisation de la charge d'un noeud

- Le modèle :

$$C(N, I_{p+1}, t_{p+1}) = C(N, I_p, t_{p+1}) + aC(vm_{p+1}, t_{p+1}) + b$$



# Critère d'évaluation

- Comment évaluer que la charge est distribuée uniformément sur les noeuds ?

- Considérer une seule variable au lieu de considérer un vecteur de charge (RAM,CPU,Net..)
  - Calculer l'écart-type

- Critère:

- Si l'écart-type est faible , la variable est uniformément distribué sur les noeud

# Critère d'évaluation

	MEM_mean	CPU_mean	netin_mean	netout_mean	MEM_std	CPU_std	netin_std	netout_std	MEM_node	CPU_Nodes	netin_node	netout_node
1616929620	2.610687	0.000000	1.747025e+05	5.073333	3.692069	0.000000	2.470666e+05	2470.666449	1=>0	0=>1	1=>0	1=>0
1616929680	2.610791	0.000000	2.210367e+05	0.526667	3.692216	0.000000	3.125931e+05	3125.930518	1=>0	0=>1	1=>0	1=>0
1616929740	2.610002	0.000000	1.885030e+05	4.283333	3.691100	0.000000	2.665835e+05	2665.834991	1=>0	0=>1	1=>0	1=>0
1616929800	2.610292	0.000000	1.562687e+05	0.000000	3.691511	0.000000	2.209973e+05	2209.972678	1=>0	0=>1	1=>0	0=>1
1616929860	2.610554	0.000000	1.791633e+05	4.283333	3.691880	0.000000	2.533752e+05	2533.752159	1=>0	0=>1	1=>0	1=>0
...	...	...	...	...	...	...	...	...	...	...	...	...
1637158860	51.204827	5.431000	1.145022e+07	5525.711667	36.216493	1.269021	1.384594e+07	138459.424998	1=>0	0=>1	0=>1	0=>1
1637158920	12.797679	3.137500	1.054849e+07	2889.175000	18.098651	4.437095	1.491782e+07	149178.211561	0=>1	0=>1	0=>1	0=>1
1637158980	12.797888	3.129167	1.044338e+07	2624.366667	18.098947	4.425310	1.476917e+07	147691.731685	0=>1	0=>1	0=>1	0=>1
1637159040	12.797976	3.115833	1.057341e+07	3028.883333	18.099071	4.406454	1.495306e+07	149530.633581	0=>1	0=>1	0=>1	0=>1
1637159100	12.797815	3.148333	6.468289e+06	4662.316667	18.098843	4.452416	9.147542e+06	91475.422646	0=>1	0=>1	0=>1	0=>1

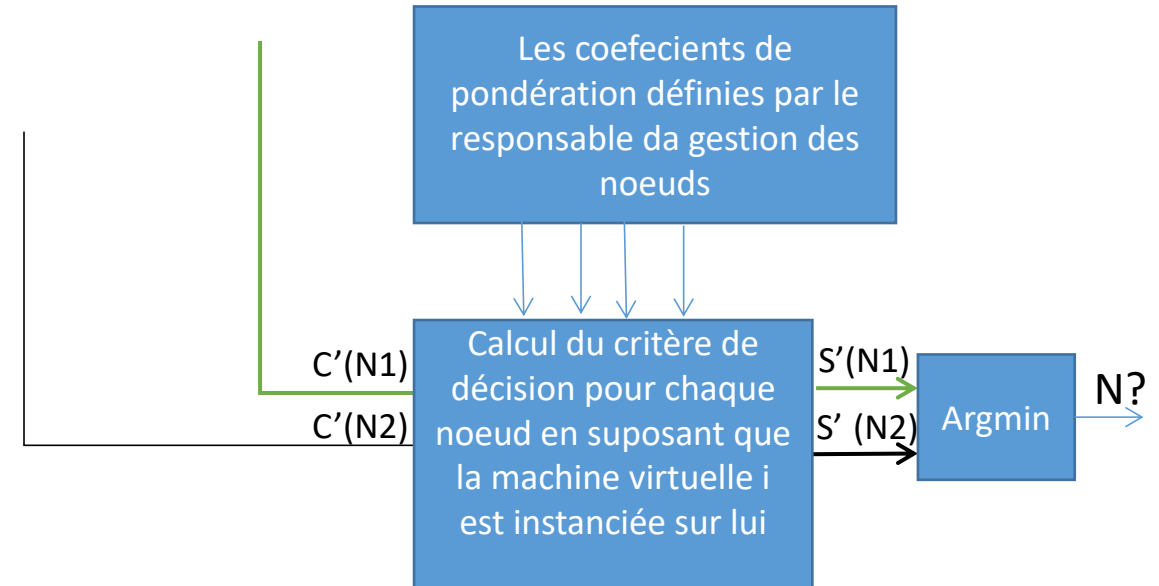
# Critère généralisé

- Equart-type normalisé

$$\sigma_v' = \frac{\sigma_v}{\max(v) - \min(v)}$$

- Moyenne pondérée

$$\sigma' = \sum_v^n \alpha_v' * \sigma_v'$$





# Algorithme de décision

---

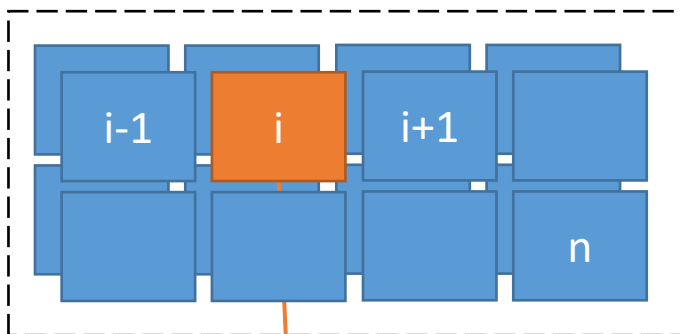
**Algorithm 1** Algorithme de décision

---

```
for  $N \in N_1, \dots, N_m$  do                                     ▷ Initialisation
     $\sigma'[N] = 0$ 
end for
 $C(vm_i) \leftarrow f_{vm_i}(t)$                                      ▷  $f_{vm_i}(t)$  c'est le modèle de prédiction 4.4.2
for  $N \in N_1, \dots, N_m$  do
     $C'(N) = C(N) + a * C(vm_i) + b$                                ▷  $C(N)$  : la charge actuelle du noeud N, voir
    l'équation (5)
     $C'(N') = C(N') \ \forall \ N' \neq N$                                ▷ C et C' sont deux vecteurs chaque composante
    représente la consommation par rapport une variable (CPU,Mémoire,...)
    for  $v \in \{CPU, NetIN, NetOUT, ..\}$  do
         $\sigma_v = \sqrt{(var(C'[v]))/(max(v) - min(v))}$ 
         $\sigma'[N] += \alpha_v * \sigma_v$                                ▷  $\alpha_v$  sont les coefficient de pondérations
    end for
end for
 $decision \leftarrow arg_{Nmin}(\sigma')$                                ▷ on choisit le noeud qui minimise  $\sigma'$ 
```

---

Les modèles des  
differentes  
machines virtuelles



Modèle de la charge d'un noeud  
 $a, b$

Noeud  
1

Noeud  
2

Les coefecients de  
pondération définies par le  
responsable da gestion des  
noeuds

Les paramètre du modèle de la  
consommation de la machine  
virtuelle i

Nouvelle machine virtuelle à  
instancier de l'utilisateur i

TimeStamp

Heure (0,..23)

Jour dans la semaine  
(Lundi , ..Dimanche)

Est-il le weekend ?  
(Oui ,Non)

Jour (1,2...31)

Prédiction de la  
consommation  
de la machine  
virtuelle i

$C(VM\_i)$

Prédiction de la charge de  
chaque noeud en suposant  
que la machine virtuelle i est  
instanciée sur lui

$a, b$

$C(N1)$

Charge actuelle de chaque  
noeud

$C(N2)$

$C'(N1)$

$C'(N2)$

Calcul du critère de  
décision pour chaque  
noeud en suposant que  
la machine virtuelle i  
est instanciée sur lui

$S'(N1)$

$S'(N2)$

Argmin

N?

# Résultats

- Évaluation du modèle de prédiction de la charge d'un noeud

$$C(N, I_{p+1}, t_{p+1}) = C(N, I_p, t_{p+1}) + aC(vm_{p+1}, t_{p+1}) + b$$

Variable	a	b
RAM	0.99585645	-545.52709379
CPU	1.85953154	-0.34058604
netIN	0.845185	-730.66786327
netOUT	0.99585645	-545.52709379

# Résultats

- Évaluation du modèle de prédiction de la charge d'un noeud

$$R = 1 - eqm(y_{pred}) / var(y_{true})$$

Variable	R
RAM	-6.5697299777050615
CPU	0.935705329122057
netIN	0.9505782370514909
netOUT	0.7888001188592341

# Résultats

- Évaluation du modèle de prédiction de la consommation d'une machine virtuelle

`err=[208, 228, 231, 242,  
159, 229, 316, 254,  
235]`

