Master Technologies de l'internet
Master Big Data
2020-2021

Cloud Computing
(Élasticité)

Meriem HALILALI
d'après des supports conçus par K.Khebbeb

meriem-sabrine.halilali@univ-pau.fr

Cloud Computing (Selon le NIST)

Définition

- *"Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model promotes availability and is composed of five essential characteristics, three service models, and four deployment models" * NIST, 2011*

UNIVERSITÉ DE PAU ET DES PAYS DE L'ADOUR

*https://www.nist.gov/sites/default/files/documents/itl/cloud/cloud-def-v15.pdf

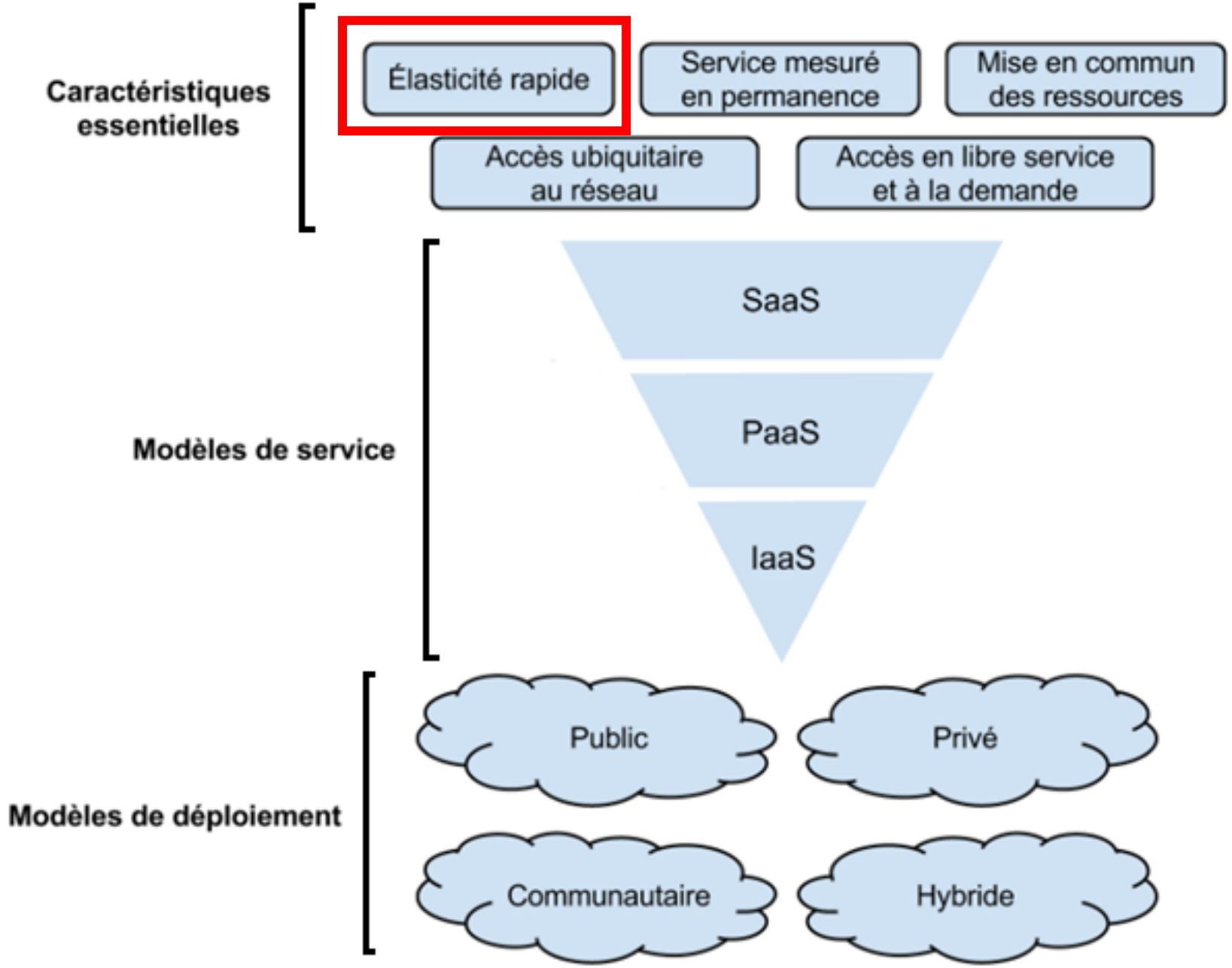
Cloud Computing (Selon le NIST)

Définition

- *"Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that* ***can be rapidly provisioned and released with minimal management effort or service provider interaction****. This cloud model promotes availability and is composed of five essential characteristics, three service models, and four deployment models" * NIST, 2011*

*https://www.nist.gov/sites/default/files/documents/itl/cloud/cloud-def-v15.pdf

Cloud Computing (Selon le NIST)



Définition de l'élasticité (1/3)

L'élasticité selon le NIST*

- *Les ressources peuvent être **facilement** provisionnées et libérées, parfois de manière automatique, afin de s'adapter à la demande (aussi bien à la hausse qu'à la baisse).*

- *Les ressources paraissent infinies et peuvent **théoriquement** être provisionnées en toute quantité et à tout moment.*

*https://www.nist.gov/sites/default/files/documents/itl/cloud/cloud-def-v15.pdf

Définition de l'élasticité (2/3)

Définition dans le monde académique

- *« L'élasticité est le degré auquel un système est capable de **s'adapter** aux **changements** de la charge de travail en ajoutant et en retirant des ressources, de manière autonomique, de sorte qu'**à tout moment**, les **ressources** disponibles **correspondent** à la **demande** actuelle aussi étroitement que possible. »** * Nikolas Herbst et al. 2013*

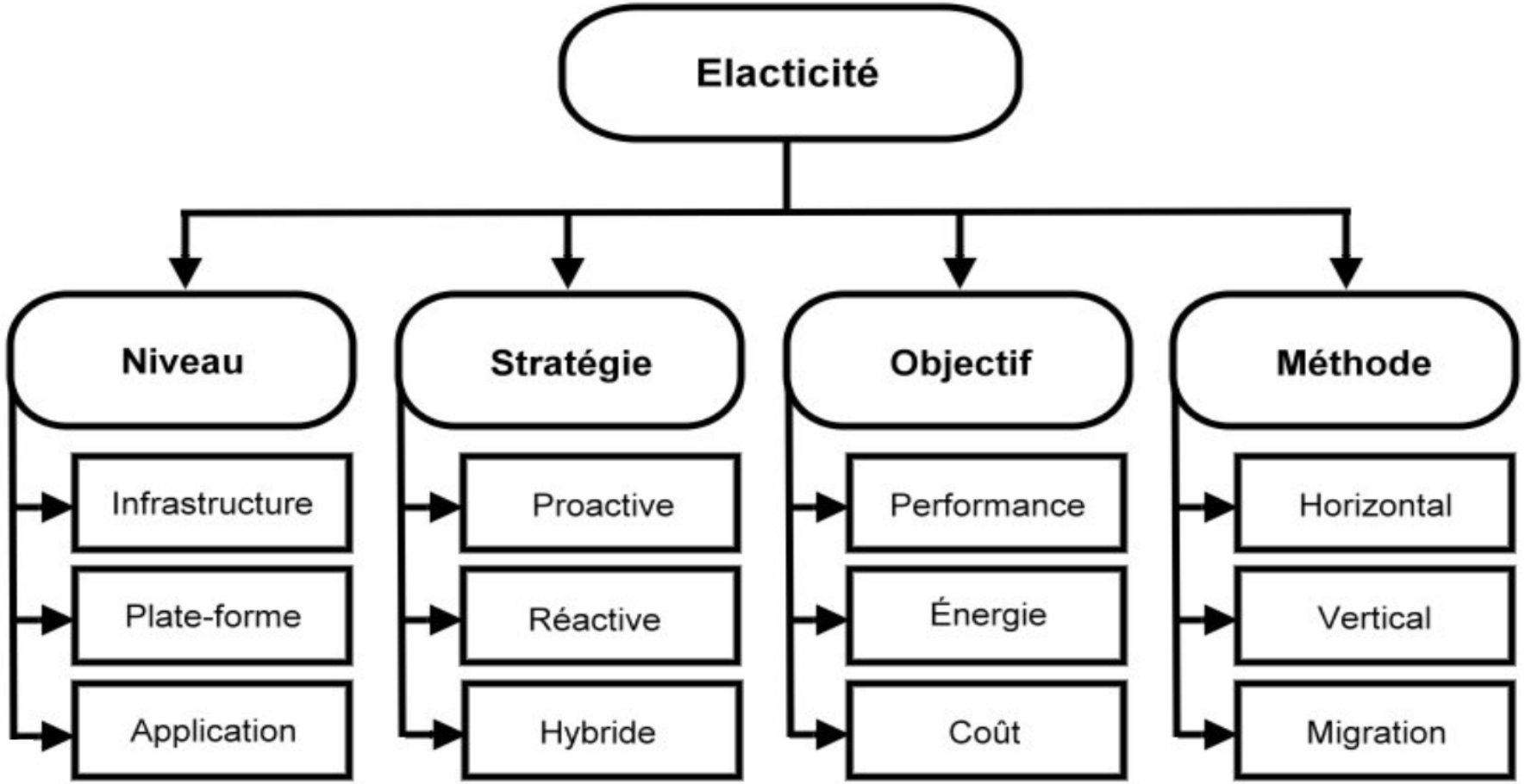
UNIVERSITÉ DE PAU ET DES PAYS DE L'ADOUR

*http://0b4af6cdc2f0c5998459-c0245c5c937c5dedcca3f1764ecc9b2f.r43.cf2.rackcdn.com/11719-icac13_herbst.pdf

Définition de l'élasticité (3/3)

Quadruplet de l'élasticité

Selon la *Classification de Galante et De Bona**



*https://www.researchgate.net/publication/235534087_A_Survey_on_Cloud_Computing_Elasticity

Objectifs de l'élasticité (1/2)

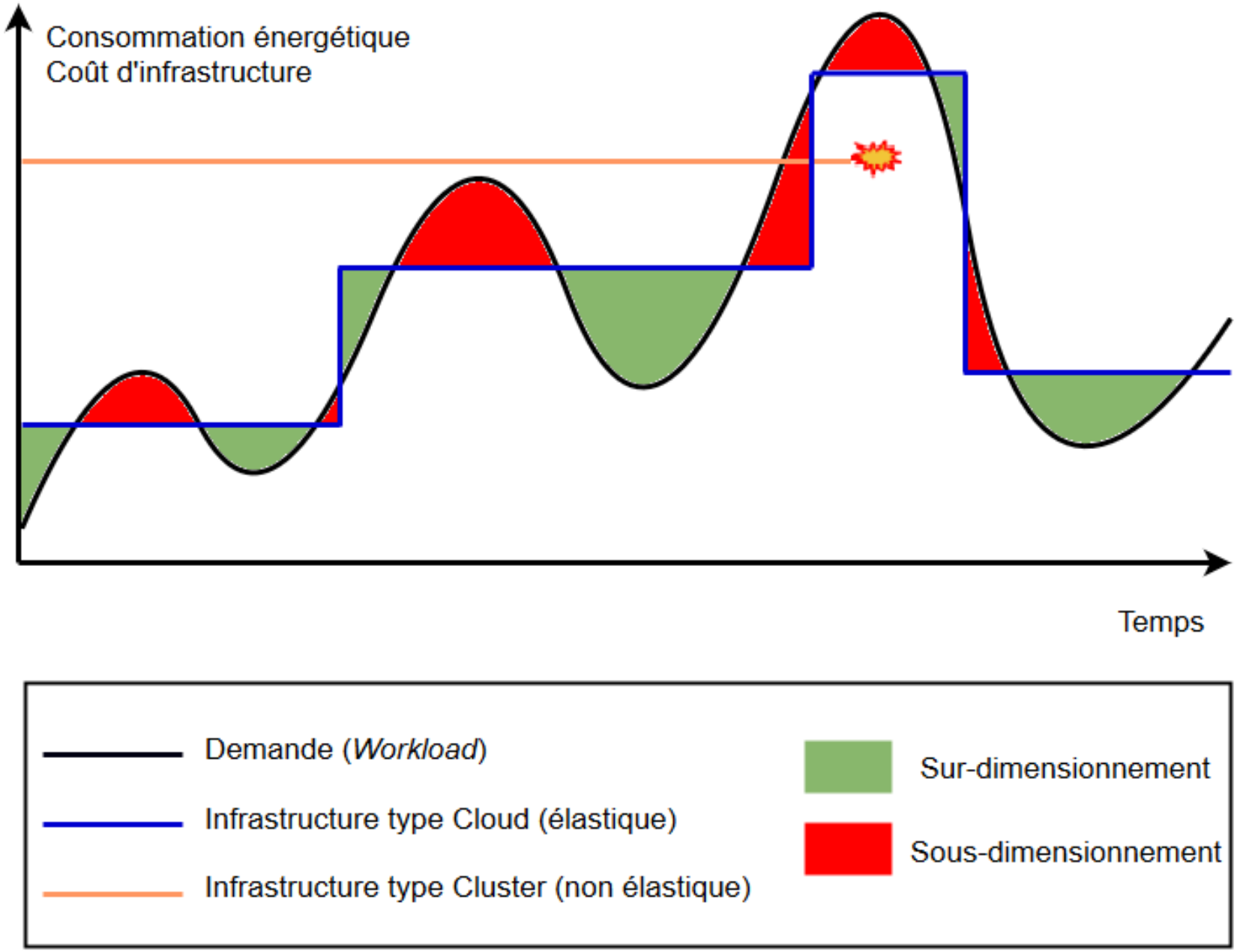
Assurer une qualité de service optimale, tout en minimisant les coûts au mieux

- *Cela revient à éviter les cas de sur-dimensionnement et de sous-dimensionnement*
- ***Sur-dimensionnement**: quantité de ressources > besoins du système*
  - *Coûts d'infrastructure importants*
- ***Sous-dimensionnement:** quantité de ressources < besoins du système*
  - *Performances médiocres*

Objectifs de l'élasticité (2/2)

Gestion des ressources

Cloud vs. Cluster



Méthodes d'élasticité (1/4)

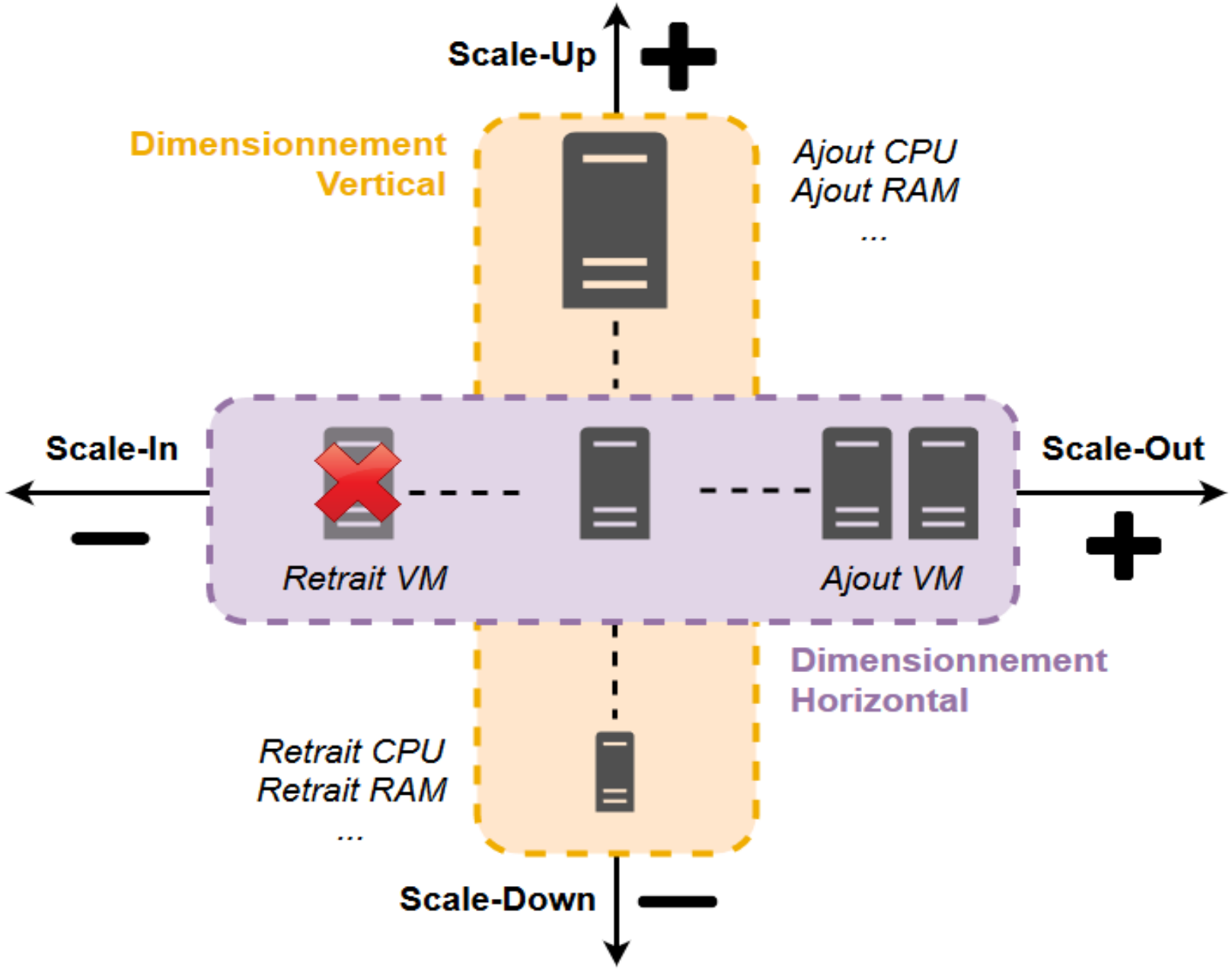
Mécanismes permettant à un système de s'adapter à sa charge de travail

- ***Dimensionnement Horizontal***
  - *Ajout (scale-out) et retrait (scale-in) de ressources (VM, service, conteneur)*
  - *Réplication / Consolidation*
  - *Étroitement lié au concept de Load-Balancing*
- ***Dimensionnement Vertical***
  - *Ajout (scale-up) et retrait (scale-down) de ressources (CPU, RAM, Stockage)*
  - *Redimensionnement*

Méthodes d'élasticité (2/4)

Gestion des ressources

au niveau Infrastructure (IaaS)



Méthodes d'élasticité (3/4)

Mécanismes permettant à un système de s'adapter à sa charge de travail

- ***Migration***
  - *Déplacer les entités (VM, Conteneur, Service) d'un hôte à un autre*
  - *Optimiser l'utilisation des ressources*
- ***Dimensionnement Hybride***
  - *Horizontal + Vertical + Migration*
  - *Compromis entre Coûts, performance, disponibilité et fiabilité*

UNIVERSITÉ DE PAU ET DES PAYS DE L'ADOUR

Méthodes d'élasticité (4/4)

Exemple Comparaison entre dimensionnement Horizontal, Vertical et Hybride*

Scaling strategy	Server scaling costs	Monitoring costs	Application's availability and reliability
Horizontal scaling	24 hrs/w × $0.085/hr × 6 servers × 52 weeks = $636.48/yr	Defining and configuring 7 metrics for 6 servers Costs: $3.5 per server/mo × 6 servers × 12 months = $252	Highly available—no single point of failure Highly likely reliable—quick recovery time
Vertical scaling	24 hrs/w × $0.68 × 1 server × 52 weeks = $848.64	Defining and configuring 7 metrics for 1 server Costs: $3.5 per server/mo × 1 server × 12 months = $42	Low availability—single point of failure Highly likely unreliable—long recovery time
Hybrid scaling	(24 hrs/w × $0.085/hr × 3 servers × 52 weeks) + (24 hrs/w × $0.34 × 1 server × 52 weeks) = $742.48	Defining and configuring 7 metrics for 4 servers Costs: $3.5 per server/mo × 4 servers × 12 months = $168	Improved availability—no single point of failure Improved reliability—medium recovery time

UNIVERSITÉ DE PAU ET DES PAYS DE L'ADOUR

*https://link.springer.com/content/pdf/10.1007%2Fs13174-011-0050-y.pdf

Stratégies d'élasticité

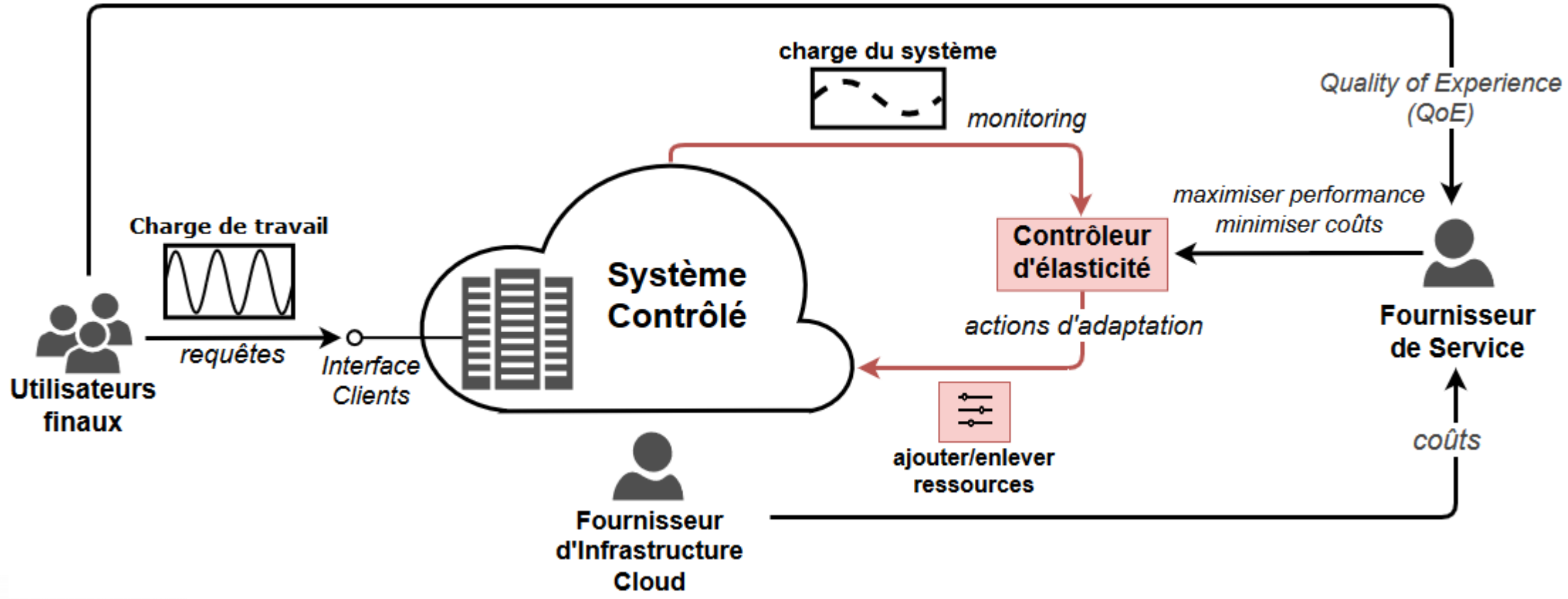
Logique gouvernant le contrôle de l'élasticité d'un système Cloud

- ***Stratégies Réactives***
  - *Règles à base de seuil (IF condition THEN action), Théorie du contrôle*
  - *Temps de réponse, quantité de ressources, demande actuelle*
- ***Stratégies Proactives (prédictives)***
  - *Théorie des files d'attente, Apprentissage par renforcement, Analyse des séries chronologiques*
  - *Demande envisagée, expériences passées, calcul de probabilités*

UNIVERSITÉ DE PAU ET DES PAYS DE L'ADOUR

Contrôleur d'élasticité (1/2)

Entité autonomique régissant le comportement élastique d'un système Cloud

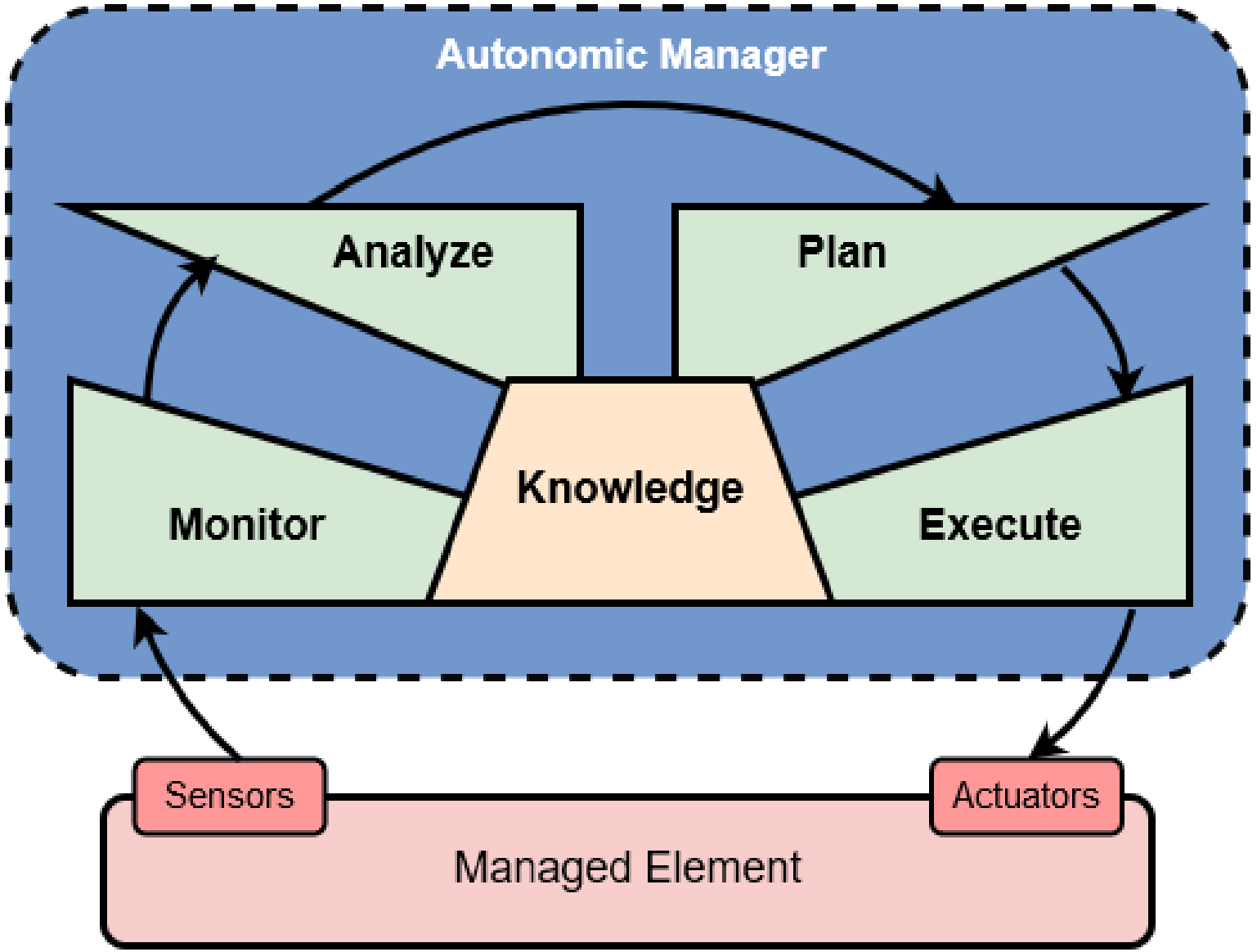


https://ieeexplore.ieee.org/document/8638257

Contrôleur d'élasticité (2/2)

Le contrôleur d'élasticité d'élasticité est un gestionnaire autonomique

- ***Boucle de contrôle Autonomique MAPE-K****



* https://www-03.ibm.com/autonomic/pdfs/AC%20Blueprint%20White%20Paper%20V7.pdf

Références

- https://cs.uwaterloo.ca/~a78khan/courses-offered/cs446/2010_05/lecture-slides/16_CloudComputing.pdf
- https://www.sanog.org/resources/sanog26/SANOG26_Tutorial%20-%20Introduction_Cloud_Computing_Sreenath.pdf
- http://web.mit.edu/smadnick/www/Courses/2013BigData/04a%20Cloud%20computing%20-%20Wikipedia.pdf
- http://csc-srv1.lasalle.edu/mccoey/csit375/elasticity.pdf