



# Placement de données dans un Cloud fédéré à base de système de stockage hybride

Amina Chikhaoui, Laurent Lemarchand, Kamel Boukhalfa & Jalil Boukhobza

Présenté par:  
**Jalil Boukhobza**

13 juin 2022

# Plan

## ① Introduction

Contexte

motivation et problématique

Architecture générale du système

## ② Approche proposée

Formulation du problème

CDP-NSGAI<sub>IR</sub>

Évaluation

## ③ Conclusion

Conclusion

# Plan

## ① Introduction

Contexte

motivation et problématique

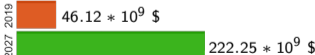
Architecture générale du système

# Contexte

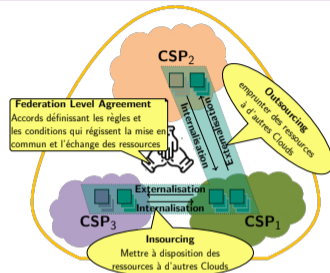
## Fédération de *Clouds*

- Collaboration de plusieurs CSP via l'internalisation et l'externalisation de la charge de travail des clients pour leur intérêt mutuel. [1, 2, 3].
- Gouvernée par un FLA.
- Plus de: performance, disponibilité, profit, équilibrage de charge.




## Stockage dans le *Cloud*

- STaaS: stockage et accès aux données via internet.
- Différents services de stockage.
- Marché STaaS [4]:  


Year	Market Size (\$)
2019	46.12 * 10 <sup>9</sup>
2027	222.25 * 10 <sup>9</sup>
- Métriques:
  - E/S: différentes technologies (HDD, SSD, etc.).
  - Latence du réseau: rapprocher les données de leurs utilisateurs.



### Avantages:

-  **Frais réduits**  
personnel, matériel, etc.
-  **Scalabilité**  
capacité nécessaire au moment opportun.
-  **Plus d'efficacité**  
intégrité, disponibilité et durabilité.

# Motivation et problématique

## Motivation

- Objectifs des fournisseurs de services *Cloud*: offrir des services attractifs aux consommateurs à des prix compétitifs.
- Le système de stockage est considéré comme l'une des options d'optimisation.
- Le *Cloud* fédéré est un environnement hautement hétérogène et dynamique.
- Une stratégie de placement doit être conçue selon l'architecture du système pour laquelle elle est proposée et les objectifs visés.

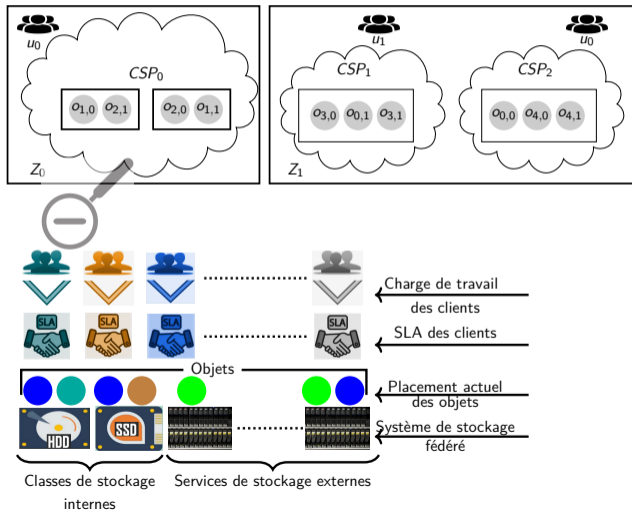
**Problématique:** Comment optimiser le placement de données pour un CSP faisant partie d'une fédération ?

- minimiser le coût global du placement pour le CSP.
- respecter les exigences des clients.

## Architecture générale du système considéré

## Hypothèses

- Fédération de plusieurs CSPs.
- Un des CSPs est concerné par l'optimisation.
- Ressources de stockage hybride:
  - 1 locales (HDD, SSD, etc.)
  - 2 services de stockage des autres CSPs.
- Utilisateurs:  
géographiquement distribués  
et / ou mobiles.



# Plan

## ② Approche proposée




Formulation du problème

CDP-NSGAI<sub>IR</sub>

Évaluation

# Formulation du problème

**Fonction objectif:** Notre modèle de Coût [5]:

-  Coût de stockage: relatif au placement futur des objets
-  Coût de latence: relatif au placement futur des objets et à la localisation future des clients.
-  Coût de migration: relatif au placement actuel et futur des objets.

## Coûts contradictoires


Exemple:


un placement local  $\implies$  peut ne pas être coûteux en termes de stockage et de migration, mais peut générer une latence élevée.




**Optimisation multi-objectifs**

## Contraintes

 Contraintes sur les bornes des capacités et performances des ressources internes et externes.

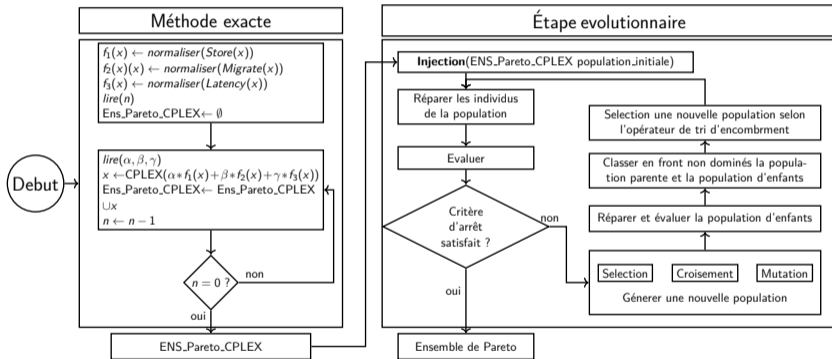
 Contraintes de SLA: les E/S offertes à chaque client doivent dépasser le seuil de dégradation de service toléré.

 Contrainte d'unicité: les objets ne sont pas répliqués.



# CDP-NSGAI<sub>IR</sub> une solution matheuristique pour le placement de données II

- CDP-NSGAI<sub>IR</sub> (a *Constraint Data Placement mathuristic based on NSGAI with Injection and Repair functions*): une matheuristique fusionnant NSGAI et une méthode exacte par programmation linéaire.



- Cette approche peut être exploitée de différentes manières:
  - **Solution 1:** seulement la méthode exacte.
  - **Solution 2 (CDP-NSGAI):** seulement NSGAI.
  - **Solution 3 (CDP-NSGAI<sub>R</sub>):** NSGAI avec la fonction de réparation.
  - **Solution 4 (CDP-NSGAI<sub>I</sub>):** NSGAI avec la fonction d'injection.
  - **Solution 5 (CDP-NSGAI<sub>IR</sub>):** NSGAI avec les fonctions d'injection et de réparation.

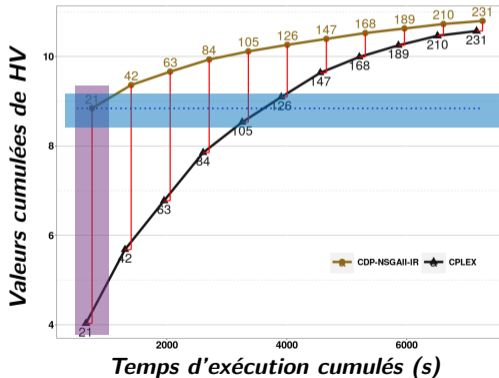
# Évaluation

## objectifs

- Montrer l'efficacité, la généralisation et la flexibilité de l'approche proposée.

## Flexibilité de l'approche proposée

- Le HV (cumulé) de CDP-NSGAIIR est toujours supérieur à celui de CPLEX.
- La matheuristique améliore CPLEX jusqu'à 2.2 fois (pour 21 solutions injectées dans la figure).
- La différence de HV est inversement proportionnelle aux nombres de solutions injectées.
- CPLEX a toujours besoin de plus de temps pour atteindre le même HV calculé par la matheuristique.



# Plan

## ③ Conclusion Conclusion

# Conclusion

## Contexte

- La fédération de Clouds permet la collaboration de plusieurs *Clouds*.
- Le placement de données dans un système de stockage fédéré est très complexe.

## Problématique

- Comment placer efficacement les objets des clients pour un CSP membre d'une fédération ?
  - ① Respecter les SLA des clients.
  - ② Minimiser le coût de placement.

## Contribution

- Formulation du problème : optimisation multi-objectifs.
- Proposition de CDP-NSGAI<sub>IR</sub> [6], une approche mathéuristique pour résoudre le problème de placement.

## Résultats

- Efficacité de l'approche proposée en termes de HV et temps d'exécution.
- CDP-NSGAI<sub>IR</sub> peut être généralisée à d'autres méta-heuristiques.

## Perspectives

- Intégration d'autres objectifs au problème de placement tels que la disponibilité et la sécurité des données.
- Intégration d'autres ressources telles que les exigences de mémoire et de charge CPU.

Merci pour votre attention

Questions?

# Références I

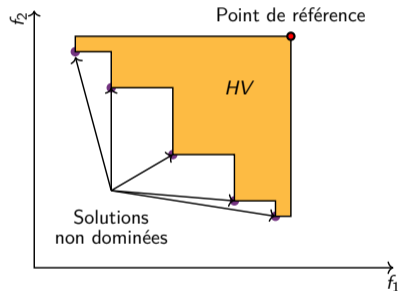
- [1] M. R. Assis and L. F. Bittencourt, "A survey on cloud federation architectures: Identifying functional and non-functional properties," *Journal of Network and Computer Applications*, vol. 72, pp. 51–71, 2016.
- [2] A. N. Toosi, R. N. Calheiros, R. K. Thulasiram, and R. Buyya, "Resource provisioning policies to increase iaas provider's profit in a federated cloud environment," in *2011 IEEE International Conference on High Performance Computing and Communications*. IEEE, 2011, pp. 279–287.
- [3] S. Rebai, M. Hadji, and D. Zeghlache, "Improving profit through cloud federation," in *2015 12th Annual IEEE Consumer Communications and Networking Conference (CCNC)*. IEEE, 2015, pp. 732–739.
- [4] "Cloud storage market," accessed February, 2021, "<https://www.alliedmarketresearch.com/cloud-storage-market>".
- [5] A. Chikhaoui, K. Boukhalfa, and J. Boukhobza, "A cost model for hybrid storage systems in a cloud federations," in *2018 Federated Conference on Computer Science and Information Systems (FedCSIS)*. IEEE, 2018, pp. 1025–1034.
- [6] A. Chikhaoui, L. Lemarchand, K. Boukhalfa, and J. Boukhobza, "Multi-objective optimization of data placement in a storage-as-a-service federated cloud," *ACM Transactions on Storage (TOS)*, vol. 17, no. 3, pp. 1–32, 2021.

## Solution multi-objectif

- Un ensemble de compromis appelés solutions non dominées.
- Critères de qualité: précision, exhaustivité et diversité.

## Métrique: Hypervolume (HV)

- Indicateur populaire de performances des algorithmes multi-objectifs.
- Hypervolume de la région multidimensionnelle fermée par un ensemble des solutions non-dominées et un point de référence.
- Plus la valeur de HV est élevée, meilleure est l'approximation.



## Efficacité de la fonction de réparation pour les petites instances

Objets	10		13		15	
	HV	Temps (s)	HV	Temps (s)	HV	Temps (s)
Méthode exacte	0.269	2.4	0.311	154	0.131	2886
CDP-NSGAI <sub>R</sub>	0.269	0.3	0.310	0.3	0.129	0.3
NSGAI	0.269	0.8	0.311	0.6	0.131	0.6
PSO	0.269	0.2	0.310	0.3	0.130	0.2

- Le temps d'exécution de la méthode exacte augmente de manière exponentielle (environ 48 minutes pour seulement 15 objets et 4 emplacements).
- Les autres algorithmes évalués obtiennent plus ou moins le même HV (différence entre 0 et 1.5%) en peu de temps (moins de 1 seconde).

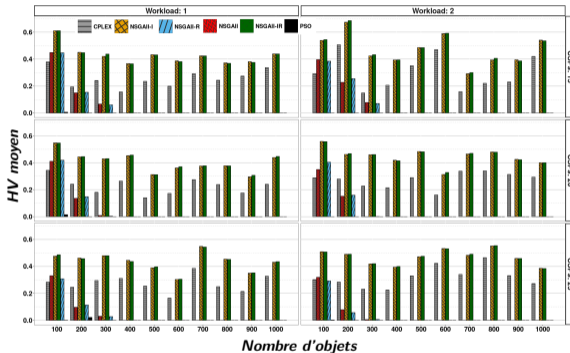


## Coefficients de la méthode exacte

$\alpha$	1	0	0	0.5	0.5	0	0.6	0.2	0.2	0.4
$\beta$	0	1	0	0.5	0	0.5	0.2	0.6	0.2	0.3
$\gamma$	0	0	1	0	0.5	0.5	0.2	0.2	0.6	0.3

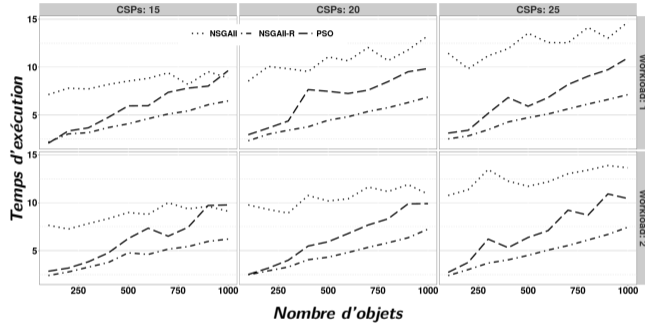
- La fonction objectif est la somme pondérée des trois coûts de stockage, migration et latence du réseau ( $\alpha * Store_N + \beta * Migrate_N + \gamma * Latency_N$ ).
- Chaque colonne correspond à une solution.

# Comparaison des HV



- CDP-NSGAI<sub>IR</sub> et CDP-NSGAI<sub>I</sub> donnent les meilleurs résultats.
- CDP-NSGAI<sub>IR</sub> améliore le HV de CPLEX de 15% à 60% et NSGAI de 4% à 94%.
- NSGAI et PSO conduisent à un HV négligeable dans la plupart des cas évalués.
- CDP-NSGAI<sub>R</sub> possède à peu près le même HV que CDP-NSGAI avec une différence de -4% à +9%.

# Évolution du temps d'exécution



- CDP-NSGAI<sub>R</sub> améliore 40% et 86% le temps d'exécution de CDP-NSGAI.
- Le temps d'exécution de CDP-NSGAI<sub>R</sub> augmente linéairement avec l'augmentation du nombre d'objets.
- En définitive, l'opérateur d'injection a un grand impact sur le HV résultant tandis que l'opérateur de réparation a un grand impact sur le temps d'exécution.

# Généralisation de la métheuristique

- La matheuristique proposée dans ce travail peut être généralisée en exploitant d'autres méta-heuristiques.
- L'ensemble des solutions calculées par CPLEX surpasse les HV de NSGAI et PSO respectivement de 46% et 57%.
- Le HV de CPLEX surpasse  $NSGAI_R$  de 44% et  $PSO_R$  de 51%.
- $NSGAI_{IR}$  et  $PSO_{IR}$  surpassent CPLEX et toutes les autres variantes d'algorithmes.

