

# Déploiement générique d'applications sur plates-formes hétérogènes réparties

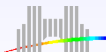
Benjamin Depardon

Sous la direction d'Eddy Caron (ÉNS-Lyon/INRIA/LIP)  
et  
Christian Pérez (IRISA/INRIA)

Juin 2007



GRAAL



PARIS  
Projet Team



# Contexte

- Calcul scientifique
  - environnements homogènes : machines parallèles
  - environnements hétérogènes répartis : Grilles
- Difficulté d'utilisation : besoin d'outils
- Problème du déploiement
  - choix des ressources
  - attribution des ressources aux applications
  - installation
  - lancement
  - configuration
- **Objectif** : proposer des heuristiques d'attribution de ressources aux applications

# Plan

- 1 Outils de déploiement
  - État de l'art
  - ADAGE
  - Extension
- 2 Heuristiques de déploiement
  - Notations
  - Problème d'optimisation
  - Hiérarchie
  - Heuristiques
- 3 Simulations

# Plan

- 1 Outils de déploiement
  - État de l'art
  - ADAGE
  - Extension
- 2 Heuristiques de déploiement
  - Notations
  - Problème d'optimisation
  - Hiérarchie
  - Heuristiques
- 3 Simulations



# Tour d'horizon des logiciels de déploiement

		ADAGE	DistAnt	GoDIET	Jade	JDF	Sekitei	Weevil
Description applications	Spécifique	x		x		x		
	Générique	x	x		x		x	x
Planning	Auto	Basique	x					
		Intelligent					x	
	Utilisateur	x	x	x	x	x		x
Compilation			x					
Installation distante		x	x		x	x		x
Pré-configuration		x	x	x	x	x		x
Post-configuration		x			x			
Lancement		x		x	x	x		x
Gestion dynamique					x			
Modularité des algorithmes		x						



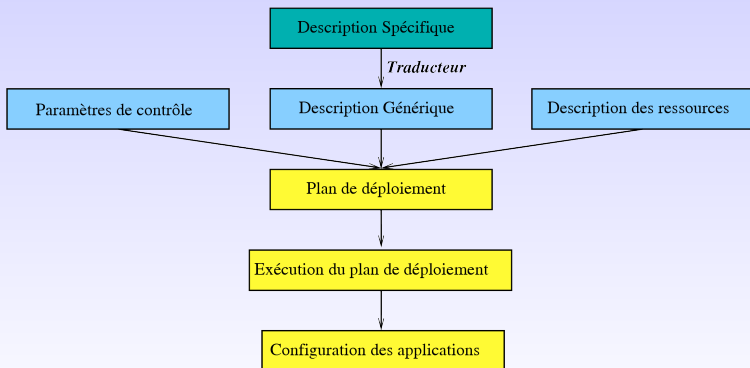
# Tour d'horizon des logiciels de déploiement

		<i>ADAGE</i>	<i>DistAnt</i>	<i>GoDIET</i>	<i>Jade</i>	<i>JDF</i>	<i>Sekitel</i>	<i>Weevil</i>
Description applications	Spécifique	x		x		x		
	Générique	x	x		x		x	x
Planning	Auto	Basique	x					
		Intelligent					x	
	Utilisateur	x	x	x	x	x		x
Compilation			x					
Installation distante		x	x		x	x		x
Pré-configuration		x	x	x	x	x		x
Post-configuration		x			x			
Lancement		x		x	x	x		x
Gestion dynamique					x			
Modularité des algorithmes		x						

# Fonctionnement général

**ADAGE** (PARIS/IRISA) :

*Automatic Deployment of Applications in a Grid Environment*



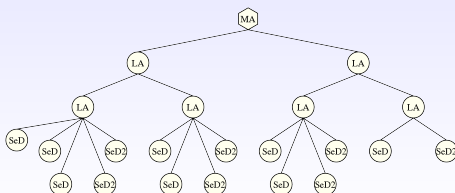
# Extensions de GADe

GADe (S. Lacour) : *Generic Application Description*

- GADe : permet de décrire presque tous les types d'applications
- Description statique

Extension

- Applications hiérarchiques
- Structure récursive
- Choix dynamique (lors du planning) de la forme de la hiérarchie





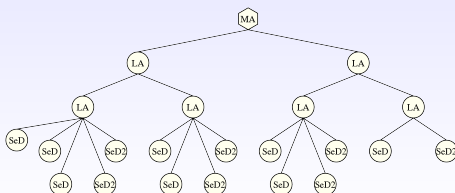
# Extensions de GADe

GADe (S. Lacour) : *Generic Application Description*

- GADe : permet de décrire presque tous les types d'applications
- Description statique

## Extension

- Applications hiérarchiques
- Structure récursive
- Choix dynamique (lors du planning) de la forme de la hiérarchie



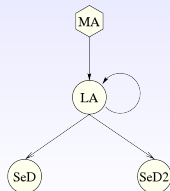
# Extensions de GADe

GADe (S. Lacour) : *Generic Application Description*

- GADe : permet de décrire presque tous les types d'applications
- Description statique

## Extension

- Applications hiérarchiques
- Structure récursive
- Choix dynamique (lors du planning) de la forme de la hiérarchie



# Plan

- 1 Outils de déploiement
  - État de l'art
  - ADAGE
  - Extension
- 2 Heuristiques de déploiement
  - Notations
  - Problème d'optimisation
  - Hiérarchie
  - Heuristiques
- 3 Simulations

## Notations

- $m$  processeurs  $p_i, i \in [1..m]$
- $n$  applications  $a_j, j \in [1..n]$
- pour chaque processeur  $p_i$  :
  - puissance de calcul  $\alpha_i$
  - coût de communication avec un autre processeur  $p_{i'} : \beta_{i,i'}$
- pour chaque application  $a_j$  :
  - nombre d'instances maximum  $nb_j^{max}$
  - besoin en travail  $w_j$
  - fréquence d'utilisation  $f_j$
  - quantité de communication vers une autre application  $a_{j'} : c_{j,j'}$
  - $\delta_i^{j,k} = \begin{cases} 1 & \text{si instance } k \text{ de } a_j \text{ placée sur } p_i \\ 0 & \text{sinon} \end{cases}$
- **Objectif** : trouver un déploiement qui répartisse la charge entre les processeurs, minimise les coûts de communications et maximise le nombre d'instances déployées

# Contraintes

## Ressources

$$\forall i, \sum_{j=1}^n \sum_{k=1}^{nb_j^{max}} \delta_i^{j,k} \cdot w_j \leq \alpha_i$$

## Instances des applications

$$\forall j, \forall k \sum_{i=1}^m \delta_i^{j,k} \leq 1 \quad \text{et} \quad \forall j, 1 \leq \sum_{i=1}^m \sum_{k=1}^{nb_j^{max}} \delta_i^{j,k} \leq nb_j^{max}$$

## Nombre d'instances déployées

$$\sum_{i=1}^m \sum_{j=1}^n \sum_{k=1}^{nb_j^{max}} \delta_i^{j,k} = TP \quad \text{et} \quad \forall j, \sum_{i=1}^m \sum_{k=1}^{nb_j^{max}} \delta_i^{j,k} \geq f_j \cdot TP$$

## Estimateur

$$LB = \frac{\sigma_{charge}^2 + comm^2}{TP}$$

# Contraintes

## Ressources

$$\forall i, \sum_{j=1}^n \sum_{k=1}^{nb_j^{\max}} \delta_i^{j,k} \cdot w_j \leq \alpha_i$$

## Instances des applications

$$\forall j, \forall k \sum_{i=1}^m \delta_i^{j,k} \leq 1 \quad \text{et} \quad \forall j, 1 \leq \sum_{i=1}^m \sum_{k=1}^{nb_j^{\max}} \delta_i^{j,k} \leq nb_j^{\max}$$

## Nombre d'instances déployées

$$\sum_{i=1}^m \sum_{j=1}^n \sum_{k=1}^{nb_j^{\max}} \delta_i^{j,k} = TP \quad \text{et} \quad \forall j, \sum_{i=1}^m \sum_{k=1}^{nb_j^{\max}} \delta_i^{j,k} \geq f_j \cdot TP$$

## Estimateur

$$LB = \frac{\sigma_{charge}^2 + comm^2}{TP}$$

# Contraintes

## Ressources

$$\forall i, \sum_{j=1}^n \sum_{k=1}^{nb_j^{max}} \delta_i^{j,k} \cdot w_j \leq \alpha_i$$

## Instances des applications

$$\forall j, \forall k \sum_{i=1}^m \delta_i^{j,k} \leq 1 \quad \text{et} \quad \forall j, 1 \leq \sum_{i=1}^m \sum_{k=1}^{nb_j^{max}} \delta_i^{j,k} \leq nb_j^{max}$$

## Nombre d'instances déployées

$$\sum_{i=1}^m \sum_{j=1}^n \sum_{k=1}^{nb_j^{max}} \delta_i^{j,k} = TP \quad \text{et} \quad \forall j, \sum_{i=1}^m \sum_{k=1}^{nb_j^{max}} \delta_i^{j,k} \geq f_j \cdot TP$$

## Estimateur

$$LB = \frac{\sigma_{charge}^2 + comm^2}{TP}$$

# Contraintes

## Ressources

$$\forall i, \sum_{j=1}^n \sum_{k=1}^{nb_j^{max}} \delta_i^{j,k} \cdot w_j \leq \alpha_i$$

## Instances des applications

$$\forall j, \forall k \sum_{i=1}^m \delta_i^{j,k} \leq 1 \quad \text{et} \quad \forall j, 1 \leq \sum_{i=1}^m \sum_{k=1}^{nb_j^{max}} \delta_i^{j,k} \leq nb_j^{max}$$

## Nombre d'instances déployées

$$\sum_{i=1}^m \sum_{j=1}^n \sum_{k=1}^{nb_j^{max}} \delta_i^{j,k} = TP \quad \text{et} \quad \forall j, \sum_{i=1}^m \sum_{k=1}^{nb_j^{max}} \delta_i^{j,k} \geq f_j \cdot TP$$

## Estimateur

$$LB = \frac{\sigma_{charge}^2 + comm^2}{TP}$$



# Programme

## Formulation du problème avec contraintes linéaires

MINIMISER  $LB$ ,

SOUS LES CONTRAINTES

$$\left\{ \begin{array}{ll} (1) & \sum_{j=1}^n \sum_{k=1}^{nb_j^{max}} \delta_i^{j,k} \cdot w_j \leq \alpha_i \quad 1 \leq i \leq m \\ (2) & \delta_i^{j,k} \in \{0, 1\} \quad 1 \leq i \leq m, 1 \leq j \leq n, 1 \leq k \leq nb_j^{max} \\ (3) & \sum_{i=1}^m \delta_i^{j,k} \leq 1 \quad 1 \leq j \leq n, 1 \leq k \leq nb_j^{max} \\ (4) & \sum_{i=1}^m \sum_{k=1}^{nb_j^{max}} \delta_i^{j,k} \geq f_j \cdot TP \quad 1 \leq j \leq n \\ (5) & 1 \leq \sum_{i=1}^m \sum_{k=1}^{nb_j^{max}} \delta_i^{j,k} \leq nb_j^{max} \quad 1 \leq j \leq n \\ (6) & \sum_{i=1}^m \sum_{j=1}^n \sum_{k=1}^{nb_j^{max}} \delta_i^{j,k} = TP \end{array} \right.$$

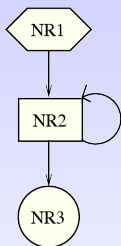
# Résultats préliminaires (P.K. Chouhan)

- Intergiciel hiérarchique
- Environnement homogène : grappe de calcul
- Régime permanent
- Objectif : maximiser le nombre de requêtes traitées par unité de temps

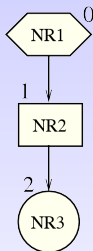
## Déploiement optimal : *CSD-Tree*

- tous les nœuds internes ont le même degré  $d$
- *sauf* peut être un à la profondeur  $h - 1$

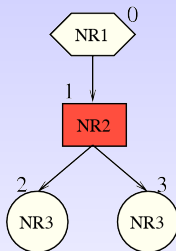
# Heuristique pour application hiérarchique



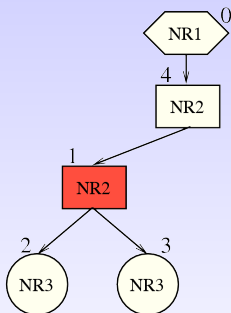
# Heuristique pour application hiérarchique



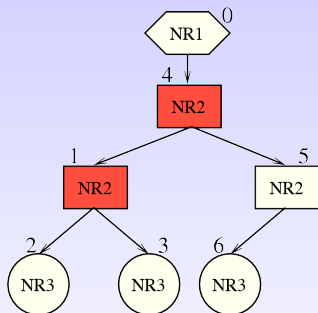
# Heuristique pour application hiérarchique



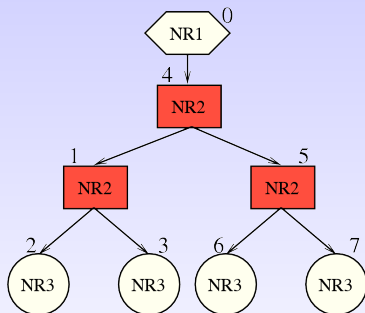
# Heuristique pour application hiérarchique



# Heuristique pour application hiérarchique

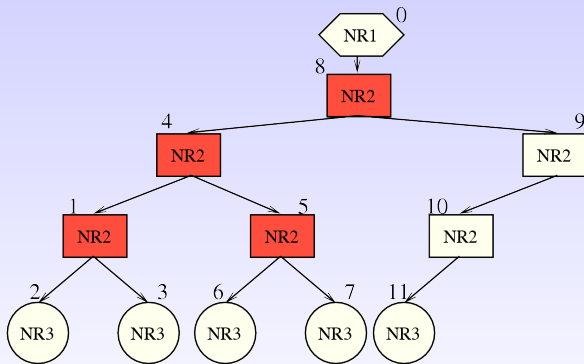


# Heuristique pour application hiérarchique



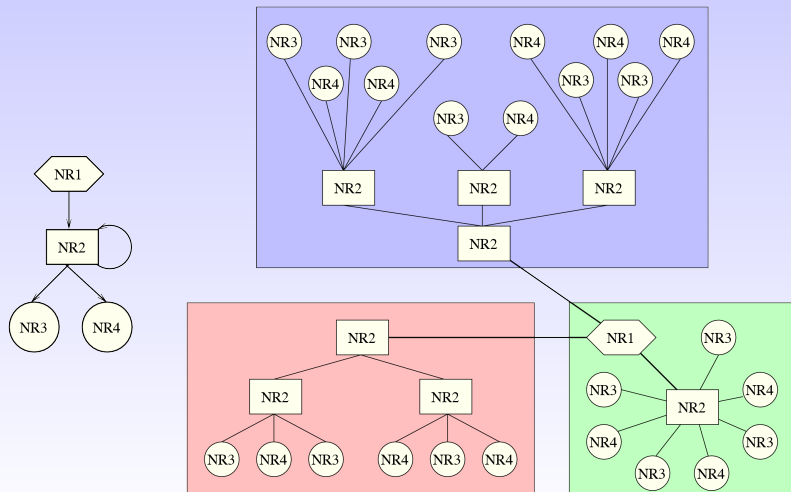


# Heuristique pour application hiérarchique



Hiérarchie

## Exemple



# Distribution processeurs, puis remplissage

## Distribution des processeurs

- Aléatoire
- Puissance de calcul
- Regroupement

## Déploiement sur les processeurs choisis

- Aléatoire
- Liste
- Rapprochement
- Glouton

# Affinité

## Première partie

- regroupement des applications communicantes
- mapping d'une instance de chaque application
- pour chaque  $a_j$  et  $p_i$ , calcul  $affinity = \sigma_{charge}^2 + comm^2$   
⇒ listes d'affinités

## Deuxième partie

Tant qu'il reste des instances à déployer

- choisir le couple  $(a_j, p_i)$  minimisant  $LB$
- mapper  $a_j$  sur  $p_i$

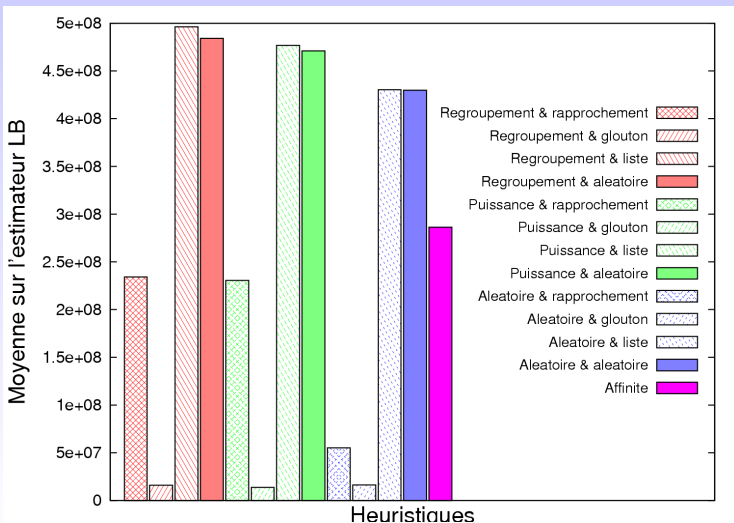
# Plan

- 1 Outils de déploiement
  - État de l'art
  - ADAGE
  - Extension
- 2 Heuristiques de déploiement
  - Notations
  - Problème d'optimisation
  - Hiérarchie
  - Heuristiques
- 3 Simulations

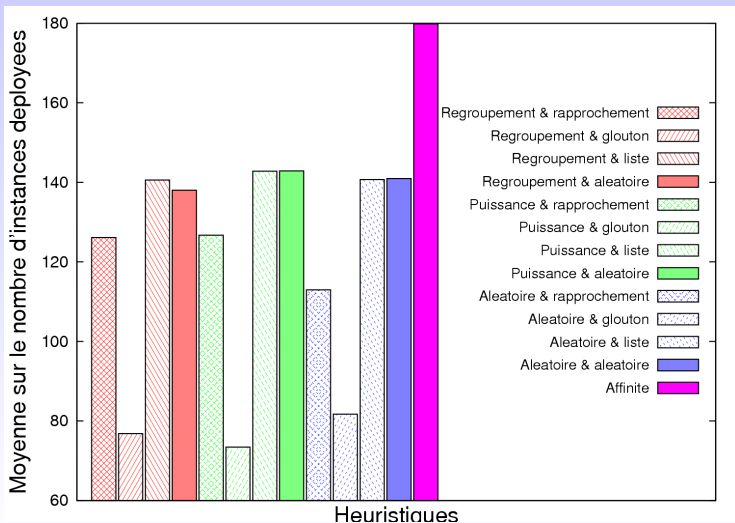
# Simulations

- Simulateur en Python
- Plate-forme
  - 4 grappes
  - 30, 20, 20 et 10 processeurs
  - 3500, 3000, 2500 et 4500 Mflops
- Graphes d'applications
  - générés à partir de 300 graines
  - nombre d'applications : entre 2 et 10
  - cardinalité max : entre 1 et  $\infty$
  - $50 \leq w_j \leq \min\{\alpha_j\}$
  - probabilité d'avoir une communication entre deux applications : 0, 10, 30, 50, 80, 100%

# Valeur de l'estimateur

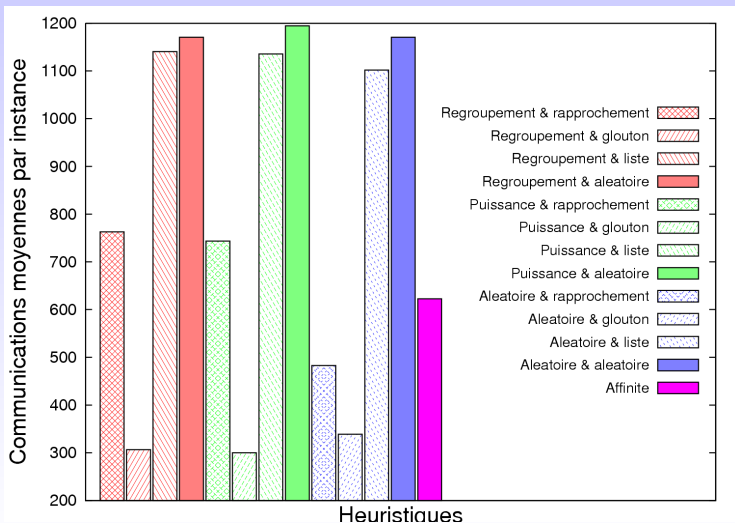


# Nombre d'instances déployées

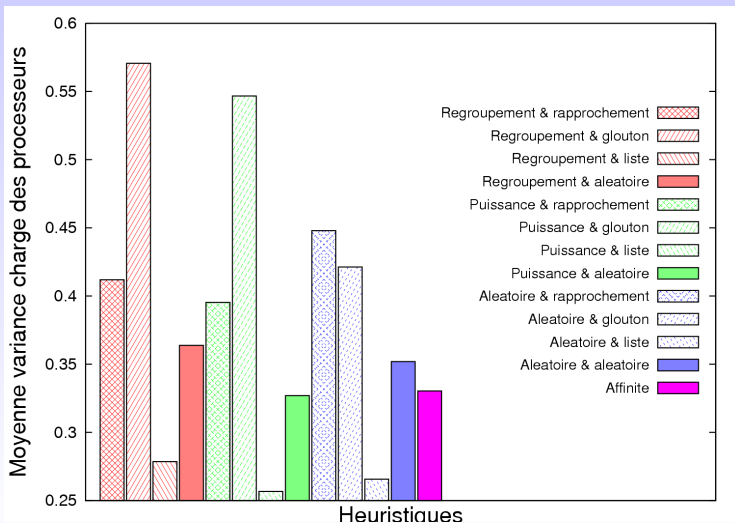




# Coût de communications par instance déployée



# Variance sur la charge des processeurs



# Résultats

- Point de vue utilisateur (nombre d'instances)
  - + Affinité
  - = Liste et aléatoire
  - – Glouton
- Point de vue administrateur
  - Coûts communications
    - + Glouton
    - = Rapprochement et affinité
    - – Liste et aléatoire
  - Charge des processeurs
    - + Liste
    - = Affinité, aléatoire et rapprochement
    - – Glouton

# Conclusion

## Contribution

- Extension du modèle de description générique d'ADAGE
- Modélisation par un problème d'optimisation avec contraintes linéaires
- Heuristique pour hiérarchies
- Comparaison de différentes heuristiques pour le cas général

## Travaux futurs

- Applications parallèles
- Redéploiement dynamique
- Implémentation au sein d'ADAGE, expérimentations sur Grid'5000
- Simplification problème d'optimisation