

ACI ReMiX

Mémoire Reconfigurable pour l'Indexation de Masses de Données

Dominique Lavenier

R2D2, IRISA, Rennes, *hardware*

S. Derrien (MCF), F. Charot (CR), G. Gorges (Ingénieur)

Symbiose, IRISA, Rennes, *bio-informatique + env. prog*

D. Lavenier (DR), F. Raimbault (MCF), S. Rubini (MCF), X. Jiang (post-doc), R. Bouville (master 2 bio-info)

TexMex, IRISA, Rennes, *image*

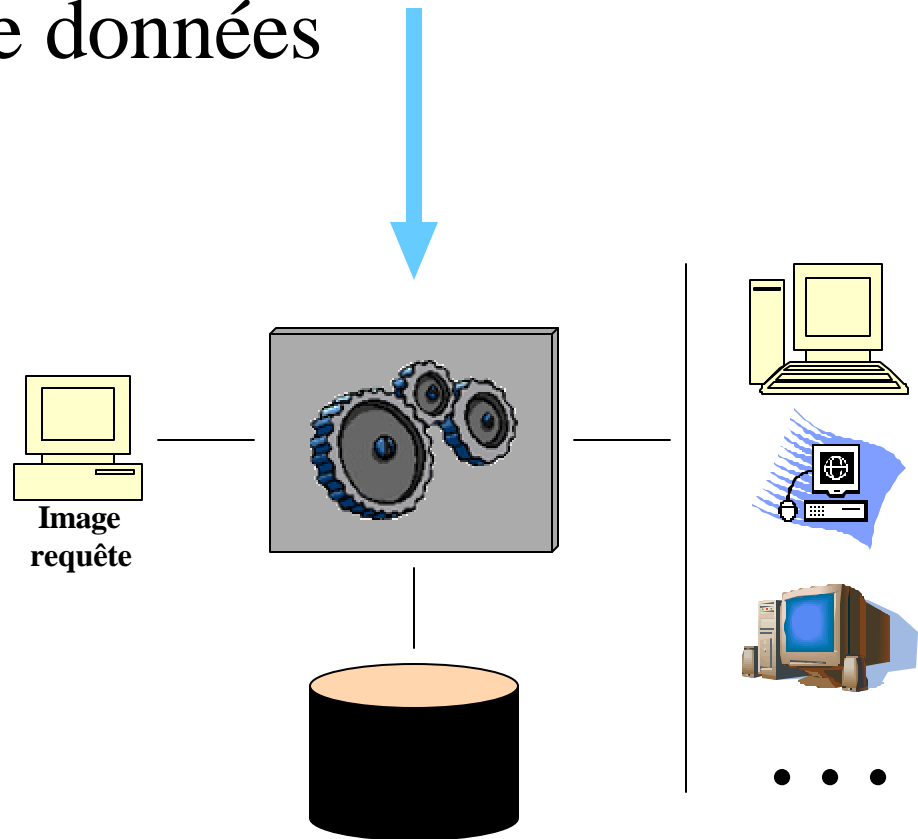
L. Amsaleg (CR), X (stage ingénieur)

Aprime, VALORIA, UBS, *textes & documents*

PF. Marteau (Prof), G. Menier (MCF), E. Popovici (Thésard)

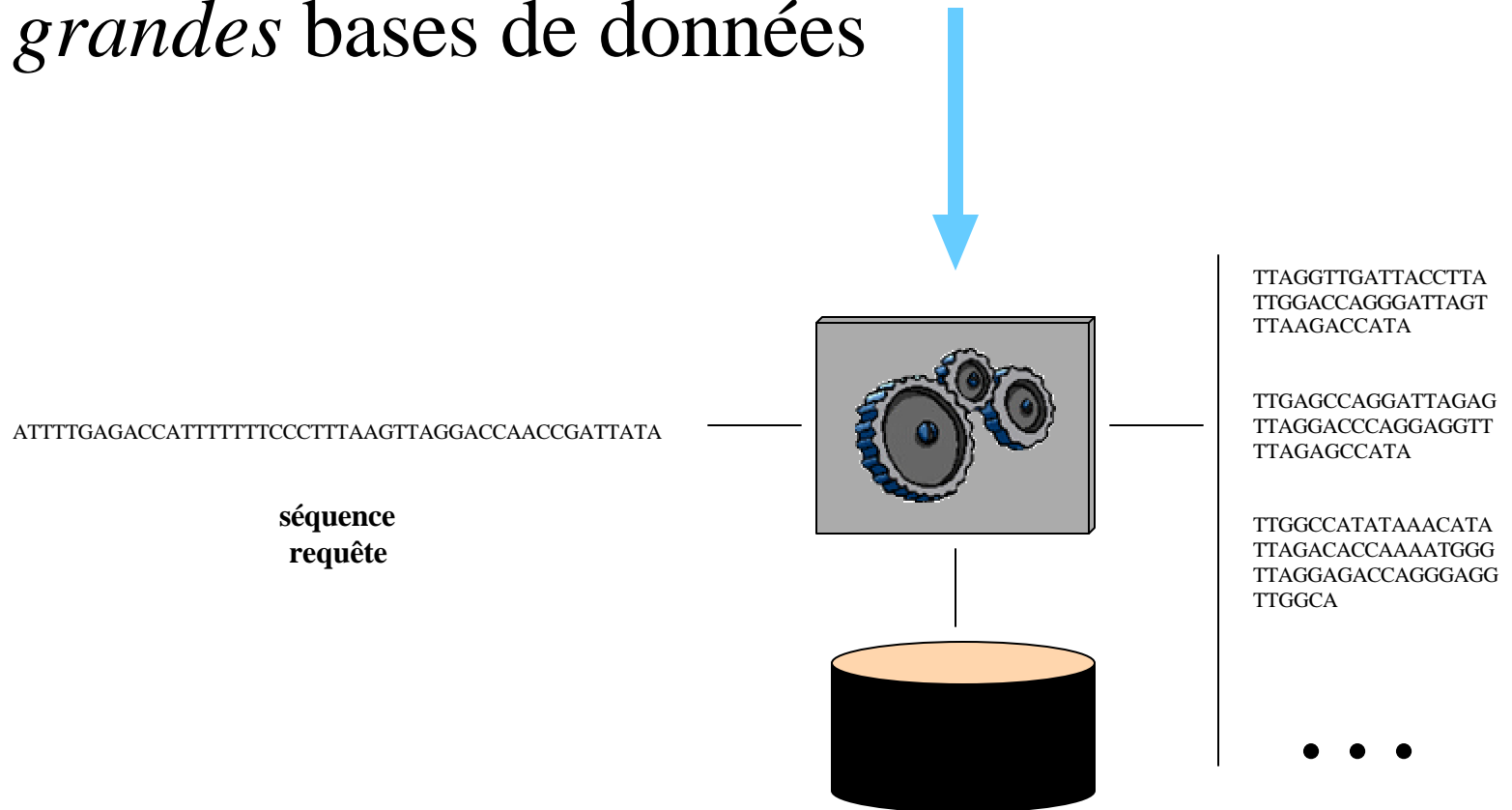
Objectifs

- Accélérer la recherche *par le contenu* dans de *grandes* bases de données



Objectifs

- Accélérer la recherche *par le contenu* dans de *grandes* bases de données



Objectifs

- Accélérer la recherche *par le contenu* dans de *grandes* bases de données



Génomique :

- $10^7 - 10^8$ séquences

Image :

- $10^6 - 10^7$ images

100 Go – 1 To

Recherche par le contenu

Solution 1 : exploration exhaustive de la base

Pour chaque objet i de la base

$d = \text{distance}(\text{objet}[i], \text{requête})$

tri (Résultat, objet[i], d)

Afficher (Résultat)

Temps de calcul :

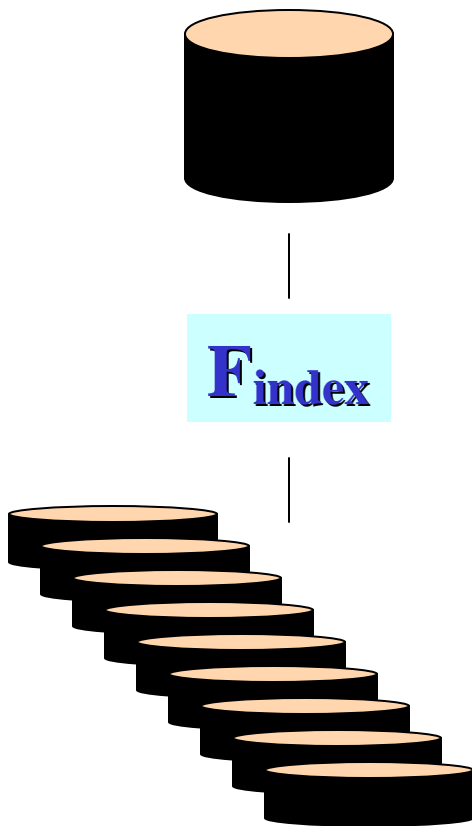
max (lecture objets , calcul distances)

Taille des données
(taux de transfert disque)

complexité du
calcul

Recherche par le contenu

Solution 2 : indexation des données de la base



liste clusters = F_{index} (requête)

Pour chaque cluster

Pour chaque objet i du cluster

$d = \text{distance}$ (objet[i], requête)

tri (Résultat , objet[i], d)

Temps de calcul :

max (lecture objets , calcul distances)

clusters
(accès disque)

complexité du
calcul

Projet ReMiX

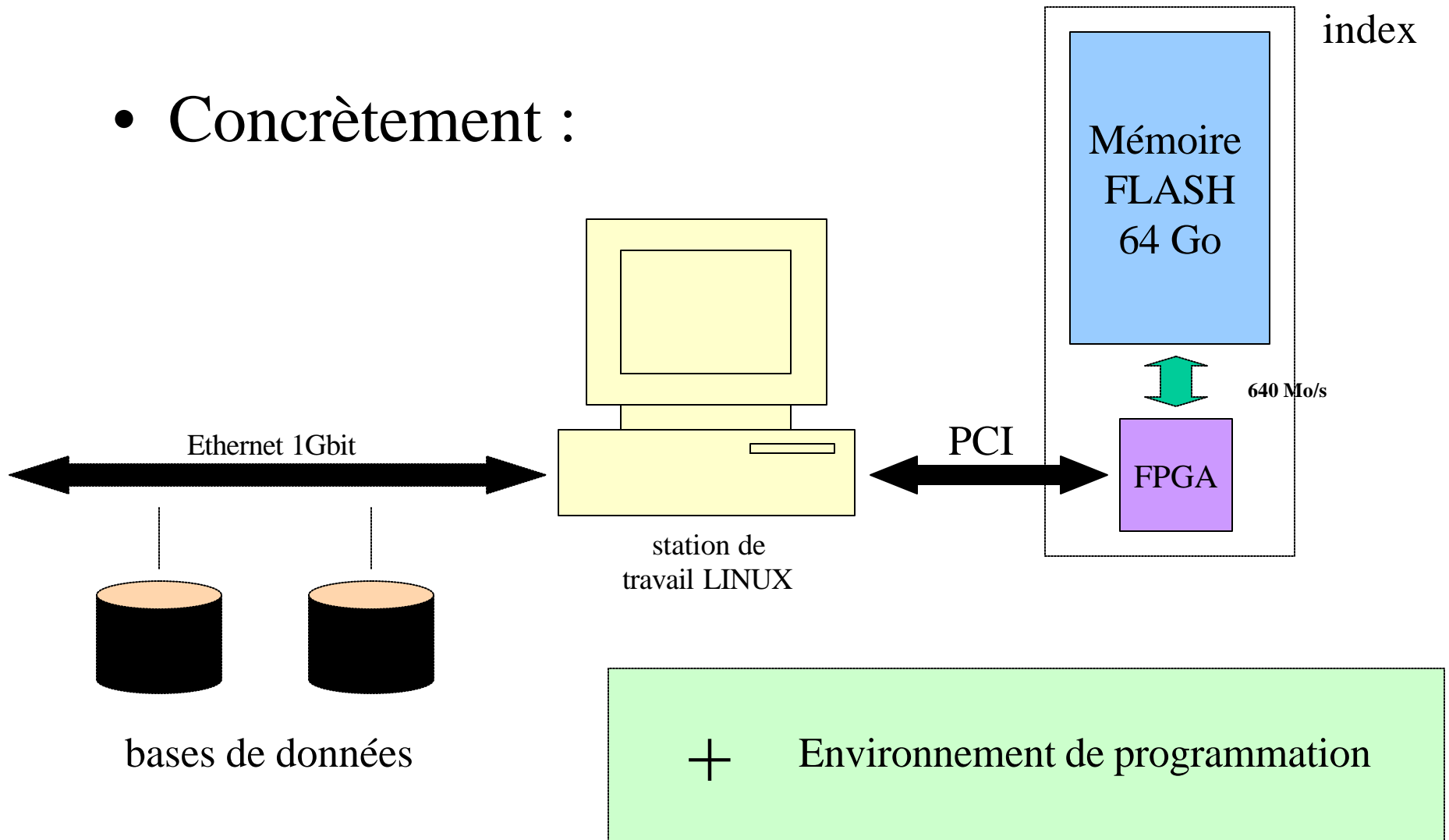
- Prototype matériel pour optimiser :
 - L'accès aux données
 - Mémoire rémanente
_____ technologie FLASH
 - Le calcul des distances
 - Opérateurs câblés
_____ technologie FPGA



Filtrage à la volée

Projet ReMiX

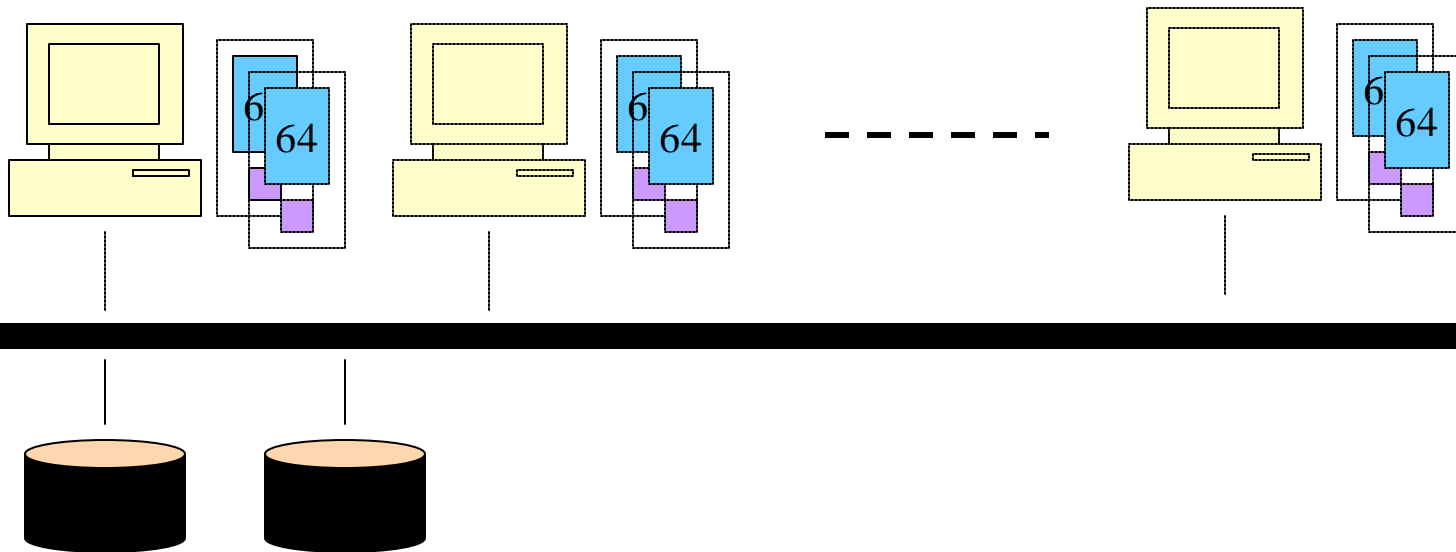
- Concrètement :



Projet ReMiX

- Augmenter la taille de l'index :

8 cartes en fabrication = 512 Go
(puce 8 Gbits)



Avancement du projet – année 1

- octobre-décembre 2003 :
 - réunion mensuel
 - mise en place de l'ACI
 - spécifications précises du projet (hard / soft)
- janvier-juillet 2004
 - développement d'un environnement de programmation
 - Framework JAVA (F. Raimbault)
 - spécification du matériel + appel d'offre
 - réflexion sur la mise en œuvre des 3 applications
 - Génomique : iBLAST
 - Image : recherche sur la base de descripteurs
 - Document XML : thèse de E. Popovici

Avancement du projet – année 2

