



Bienvenue

Portes Ouvertes sur les compétences
du Laboratoire Bordelais
de Recherche en Informatique
Jeudi 28 février 2013





Gérer le capital données de l'entreprise





Compétences LaBRI

- Visualisation de données
- Optimisation du calcul distribué
- Image/Son
- Calcul haute performance
- Fouille de données
- Graphes
- Combinatoire
- Logique





Exemples de données/projets

- Analyse de réseaux complexes
 - Réseaux sociaux
 - Internet
 - Réseaux biologiques
- Analyse de données multi-dimensionnelles
 - Données astronomiques (détection d'exoplanètes)
 - Données environnementales (prédiction climatique)





Caractéristiques

- Données massives (Big Data)
 - Problème: calcul consommateur de
 - Mémoire
 - Temps
 - Exploiter, interagir, ère du tout immédiat





Outils

- **Pré-calcul**: pré-calculer le moins possible pour atteindre un seuil de performance fixé
- **Paralléliser** les traitements pour accélérer les performances → bon **ordonnement**
- Données **distribuées**: tenir compte des coûts de **communication** → bon ordonnancement
- Algorithmes **approchés**: Quelles **garanties**?





Outils

- **Echantillonner**: quand est-ce que c'est possible?
- Données en **flux**: obtenir le résultat (approché) en une seule passe
- **Visualisation** interactive → choisir ce qui est pertinent à montrer et comment le montrer
- **Logique**: Qu'est-il possible d'exprimer et avec quelle complexité ?





Ex: Matérialisation de cubes de données

Technique

	exacte	approchée
complet	<i>24h/100Go</i>	<i>1h/105Go</i>
partiel	<i>25mn/106Go</i>	<i>2mn/110Go</i>

Espace de recherche





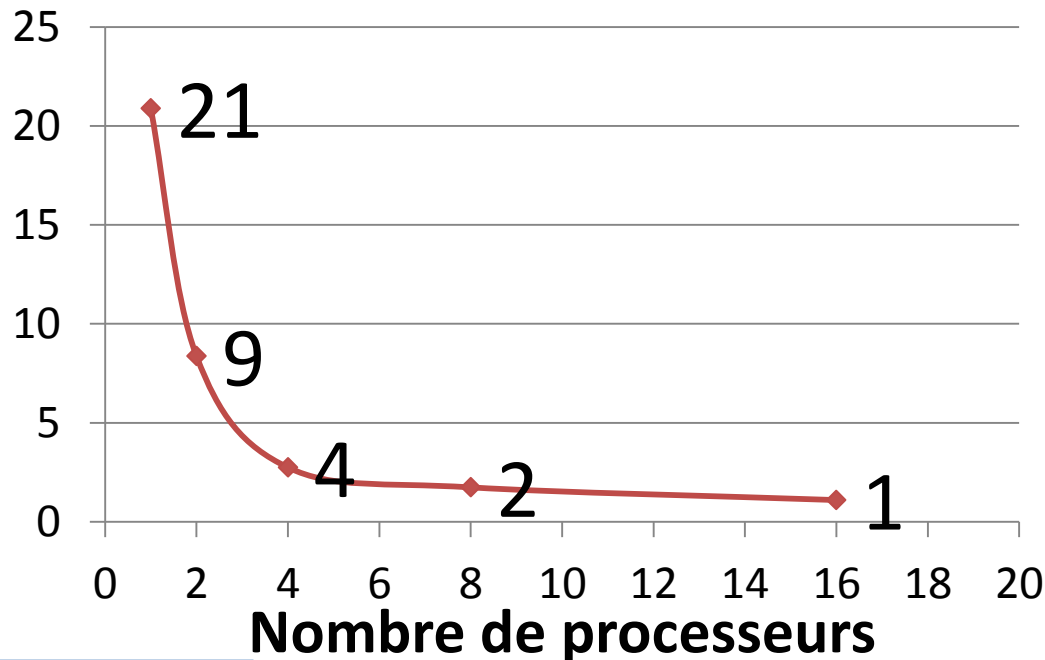
Ex: Recherche // des fréquents maximaux

1.7^{E6} transactions (fichiers téléchargés à partir d'une adresse IP), 5.3^{E6} fichiers, Moyenne: 178 fichiers/IP.

Fichiers fréquemment téléchargés par une même IP ?

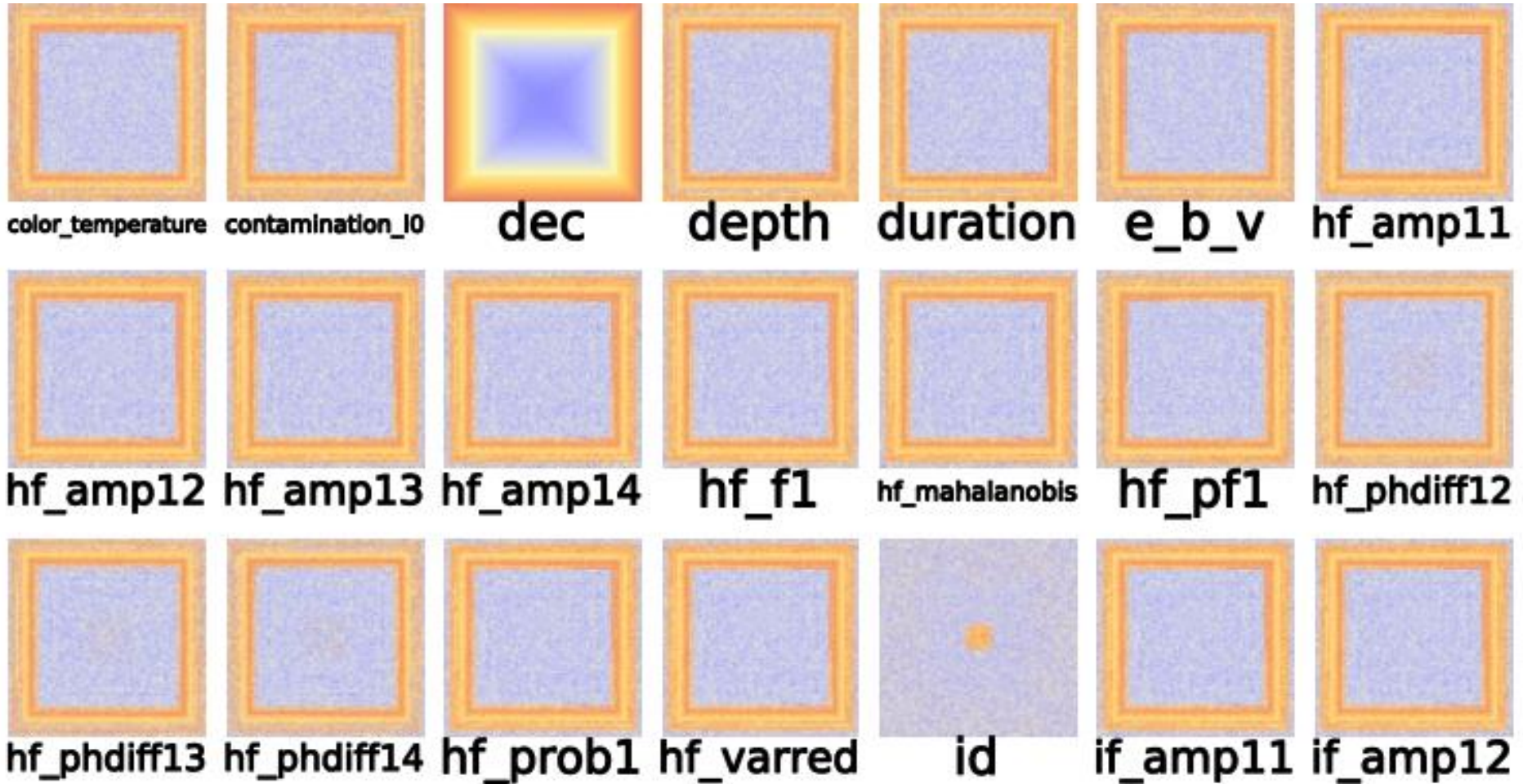
Fréquent = au moins 10% des transactions

Temps (min.)





Ex: Visualisation de données en astronomie





Ex: collaboration industrielle

- A travers une thèse CIFRE
- PME: société 2MORO
- Sujet: exploiter les données historiques d'une flotte de véhicules pour anticiper les opérations de maintenance
- Problèmes algorithmiques, approximation, résumé, pertinence, optimisation





Quelques thèses

- Optimisation de requêtes distribuées dans un système intégré
- Méthodes de comparaison de réseaux métaboliques en bio-informatique
- Analyse symbolique de données multi-génomique dans un contexte RDF distribué
- Modèle et Méthode pour la visualisation et l'analyse de données multi-dimensionnelles nominatives dynamiques
- Visualisation interactive de graphes dynamiques

