



ENTREPÔTS, REPRÉSENTATION
& INGÉNIERIE des CONNAISSANCES



EDA - 2015

11ème journées
francophones sur les
Entrepôts de Données et
l'Analyse en ligne

Rado RATSIMBAZAFY

Sous la direction de :

Omar BOUSSAID

Fadila BENTAYEB

D-Worm: Data warehouse
workload Manager :
Traitement des requêtes massives



UNIVERSITÉ
LUMIÈRE
LYON 2
UNIVERSITÉ DE LYON





Sommaire

- Contexte et problématique
- Notre approche
- Déroulement de la démonstration
- Perspectives



Contexte et problématique

- Avènement des big data
- Profusion des données ouvertes
- Evolution des supports mobiles
- Analyse en ligne dans le cloud
- Apparition des **requêtes massives**





Notre approche

Phase 1 : à partir d'une charge d'apprentissage

- Déterminer à l'aide d'un algorithme de clustering le regroupement des requêtes
- A partir de chaque cluster de requêtes, construire la requête représentative
- Construire les vues matérialisées des requêtes représentatives



Notre approche

Phase 2 : Répondre à une charge

- Pour chaque requête de la charge
 - si la classe de la requête est connue → utiliser la vue matérialisée créée précédemment
 - sinon retour à la phase 1 pour les requêtes non classifiées
- Execution parallèle des requêtes
- Récupération des données non présentes dans la vue matérialisée
- Amélioration des clusters de requêtes



Démonstration

- Regroupement des requêtes
 - Transformation des requêtes en matrices binaires
 - Execution de l'algorithme PROXIMUS
- Construction des requêtes représentatives
 - Sélection des attributs les plus représentatifs pour chaque requête représentative
- Execution d'une charge



Perspectives

- Réalisation des expérimentations dans un cloud
- Réduction du coût de stockage des vues matérialisées
- Amélioration des clusters de requêtes

MERCI