

# Modèles de Coût pour la Sélection de Vues Matérialisées dans le Nuage, Application aux Services Amazon EC2 et S3

Romain Perriot\*, Jérémy Pfeifer\*, Laurent d'Orazio\*,  
Bruno Bachelet\*, Sandro Bimonte\*\*, Jérôme Darmont\*\*\*

\*Clermont Université, CNRS, Université Blaise Pascal, LIMOS UMR 6158  
nom.prenom@univ-bpclermont.fr

\*\*IRSTEA, UR TSCE, Clermont-Ferrand  
sandro.bimonte@irstea.fr

\*\*\*ERIC Lyon 2, Université de Lyon  
jerome.darmont@univ-lyon2.fr

**Résumé.** La performance des entrepôts de données est classiquement assurée grâce à des structures comme les index ou les vues matérialisées. Dans ce contexte, des modèles de coût permettent de sélectionner un ensemble pertinent de ce type de structures. Toutefois, cette sélection devient plus complexe dans les nuages informatiques, car en plus des temps de réponse, il faut simultanément optimiser le coût monétaire. Nous proposons dans cet article de nouveaux modèles de coût intégrant le paiement à la demande en vigueur dans les nuages. Sur la base de ces modèles, nous définissons un problème d'optimisation consistant à sélectionner, parmi des vues candidates, celles à matérialiser pour minimiser le coût d'interrogation et de maintenance, ainsi que le temps de réponse pour une charge de requêtes donnée. Dans un premier temps, nous optimisons les deux critères séparément : le temps est optimisé sous contrainte de coût et vice versa. Notre proposition est ensuite validée de manière expérimentale.

## 1 Introduction

L'informatique dans le nuage (*cloud computing*), sous l'impulsion d'entreprises comme Google, Microsoft et Amazon, suscite une attention particulière. Ce paradigme permet un accès à la demande à des ressources configurables pouvant être rapidement mises à disposition avec une maintenance minimale. Selon un modèle de facturation à l'utilisation, les clients ne payent que pour les ressources (stockage et calcul) qu'ils utilisent. Les performances dans les nuages reposent généralement sur l'utilisation d'un grand nombre d'instances avec un parallélisme transparent pour l'utilisateur.

Les entrepôts de données et les systèmes OLAP (On-Line Analytical Processing) représentent des technologies d'aide à la décision qui permettent l'analyse en ligne de gros volumes de données. Ces technologies reposent sur des techniques d'optimisation comme des index, des caches ou encore des modèles logiques dénormalisés qui permettent l'analyse multidimensionnelle (agrégations sur plusieurs axes d'analyse) des données tout en garantissant des bonnes