

Intégration des Tableaux Multidimensionnels en Pig pour l'Entreposage de Données sur les Nuages

Laurent d'Orazio*, Sandro Bimonte**

*Université Blaise Pascal - LIMOS

Campus des Cézeaux

63173 Aubière Cedex, France

dorazio@isima.fr

<http://www.isima.fr/dorazio/>

**Cemagref

Campus des Cézeaux

63173 Aubière Cedex, France

sandro.bimonte@cemagref.fr

<http://eric.univ-lyon2.fr/sbimonte/>

Résumé. Les entrepôts de données et les systèmes OLAP correspondent à des technologies d'aide à la décision. Ils permettent d'analyser à la volée de gros volumes de données représentés en fonction d'un modèle multidimensionnel. L'informatique dans les nuages, sous l'impulsion des grandes compagnies telles que Google, Microsoft ou encore Amazon, a récemment suscité une attention particulière. Considérer l'interrogation OLAP et les entrepôts de données au sein de telles infrastructures devient alors un enjeu majeur. Les problèmes devant être considérés sont ceux classiques des systèmes largement distribués (interrogation de gros volumes de données, hétérogénéité sémantique et structurelle ou encore variabilité), mais d'un nouveau point de vue devant considérer les spécificités de ces architectures (facturation à l'utilisation, élasticité et facilité d'utilisation). Dans ce papier nous abordons dans un premier temps les règles de facturation à l'utilisation pour le stockage des entrepôts de données. Nous proposons d'utiliser des techniques de stockage pour nuages à base de tableaux multidimensionnels. De premières expérimentations montrent l'intérêt de notre proposition. Ensuite, nous listons des perspectives de recherche.

1 Introduction

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 [17]. Dans une architecture Relational OLAP (ROLAP), les systèmes OLAP sont déployés en utilisant des Systèmes de Gestion de Bases de Données relationnels pour stocker et analyser les données. Cette approche est pertinente pour les entrepôts de données peu denses. Quand les données sont denses, l'approche Multidimensional OLAP (MOLAP) peut être utilisée [34]. L'architecture MOLAP stocke les données à l'aide de structures multidimensionnelles