
Une approche pour l'extension à la demande de cubes multidimensionnels dans un contexte multi-modèles : Application à l'agroécologie basée sur l'internet des objets

Sandro Bimonte¹, Fagnine Alassane Coulibaly¹, Stefano Rizzi²

1. TSCF-INRAE, Université de Clermont Auvergne
9 avenue Blaise Pascal, 63170 Aubière, France
prenom1-prenom2.nom@inrae.fr

2. DISI, Université de Bologne
2 Viale del Risorgimento, 40136 Bologna, Italie
prenom.nom@unibo.it

Cet article est une synthèse de l'article : Sandro Bimonte, Fagnine Alassane Coulibaly, Stefano Rizzi. An approach to on-demand extension of multidimensional cubes in multi-model settings: Application to IoT-based agro-ecology, Data & Knowledge Engineering, Volume 150, 2024, 102267, ISSN 0169-023X, <https://doi.org/10.1016/j.datak.2023.102267>

La gestion, l'intégration et l'analyse des données semi et non structurées représentent des défis majeurs dans les écosystèmes de données. Ces systèmes doivent également être capables d'évoluer de manière transparente et efficace. Cette problématique est d'autant plus pertinente pour les analyses de type OLAP, qui exploitent des cubes multidimensionnels au sein des entrepôts de données (data warehouse, DW). Les DW stockent en effet une large variété de données, en s'appuyant sur divers modèles de données. Si la question de la gestion de cette diversité de modèles, via l'utilisation d'entrepôts de données multi-modèles (Bimonte et al., 2022), a été partiellement abordée, les enjeux liés à l'extensibilité des systèmes restent peu explorés.

L'extensibilité se réfère à la capacité d'intégrer de nouveaux éléments multidimensionnels dans les entrepôts de données multi-modèles (multi-models data warehouse, MMDW), qui n'étaient pas prévus ou identifiés lors de leur conception initiale, afin de les rendre disponibles pour des analyses OLAP ultérieures. Cela représente un avantage clé des MMDW par rapport aux DW purement relationnel. Cette caractéristique est particulièrement importante dans le contexte de l'émergence de nouvelles technologies de collecte de données issue de l'Internet des Objets (IoT).

Dans (Coulibaly et al., 2023), nous avons exploré une méthode pour implémenter l'extensibilité des cubes multidimensionnels, en adoptant une approche de définition de schéma lors de la lecture des données (Schema-on-Read). Cette technique se distingue de l'approche conventionnelle de schéma à l'écriture (Schema-on-Write), car elle permet de spécifier le schéma des données au moment où elles sont lues, au lieu de le faire lors de leur enregistrement. Une architecture a été proposée pour mettre en œuvre cette approche.

Cet article s'inscrit dans la continuité du précédent et présente « xCube », une méthode innovante favorisant l'extension dynamique des cubes multidimensionnels en réponse aux nouveaux besoins d'analyse et d'intégration de nouvelles données par les utilisateurs. « xCube » offre aux utilisateurs la possibilité de sélectionner un composant multidimensionnel (tel qu'un fait ou un niveau hiérarchique) à étendre avec des données additionnelles. Ces données additionnelles peuvent provenir soit d'un lac de données ou soit être déjà incluses dans le MMDW. Le processus d'extension du schéma multidimensionnel prend en compte les dépendances fonctionnelles entre ces nouvelles données et l'élément à étendre. En effet, ce processus d'extension utilise trois algorithmes spécifiques au type d'extension souhaité qu'il s'agisse de l'ajout de nouvelles mesures, dimensions ou niveaux hiérarchiques. Ces trois algorithmes s'appuient sur un quatrième algorithme qui renvoie l'ensemble des sommets candidats pour étendre l'élément du MMDW sélectionné. Cet algorithme recherche les dépendances fonctionnelles entre l'élément à étendre et les données additionnelles en parcourant les schémas relationnels, de documents et de graphes ainsi que les références inter-modèles. L'exécution de tous ces algorithmes permet d'obtenir un nouveau schéma multidimensionnel étendu. Ce schéma étendu est ensuite rendu accessible aux utilisateurs pour l'analyse OLAP.

Par ailleurs, nous avons proposé une implémentation de preuve de concept pour xCube dans le contexte de la surveillance épidémiologique de la Flavescence Dorée dans les vignobles de la région de Bordeaux. Nous avons utilisé AgensGraph pour le stockage de données multi-modèles et Mondrian comme serveur OLAP. Nous avons montré que l'utilisation des balises « View » et « MeasureExpression » de Mondrian permet d'utiliser comme dimensions, niveaux et mesures des données de type document et graphe stockées dans AgensGraph en encapsulant les requêtes de documents et de graphes dans des requêtes SQL. De cette manière, la complexité du stockage multi-modèle est transparente pour le serveur OLAP et, surtout, pour le client OLAP.

Bibliographie

- Bimonte, S., Gallinucci, E., Marcel, P., Rizzi, S. (2022). Data variety, come as you are in multi-model data warehouses. *Information Systems* 104, 101734.
<https://doi.org/10.1016/j.is.2021.101734>
- Coulibaly, F.A., Bimonte, S., Rizzi, S., Malembic-Maher, S., Fabre, F. (2023). Towards a Multi-Model Approach to Support User-Driven Extensibility in Data Warehouses: Agro-ecology Case Study. *EDBT/ICDT Workshops* 2023.