

Linked Open Data pour les Entrepôts de Données: Opportunité et Défis

Nabila Berkani*, Selma Khouri*, Ladjel Bellatreche**

*Laboratoire LCSi, Ecole Nationale Supérieure d'Informatique, Alger, Algérie
(n_berkani, s_khouri)@esi.dz

**LIAS/ISAE-ENSMA, 86960 Futuroscope, France
bellatreche@ensma.fr

Résumé. De nos jours, nous vivons dans un monde ouvert et connecté, où les petites, moyennes et grandes entreprises cherchent à s'approprier des systèmes de stockage pour une meilleure analyse des données. Les entrepôts de données (\mathcal{ED}) sont un exemple de ces systèmes, qui ont connu dernièrement une baisse de régime avec l'apparition des données massives et leurs différents V (Volume, Variété, Vitesse, Valeur, etc). L'arrivée de l'ère Linked Open Data (\mathcal{LOD}) représente une excellente opportunité pour les \mathcal{ED} , car elles sont porteuses de Valeur ajoutée. La prise en compte de cette Valeur dans la conception des \mathcal{ED} doit faire face au problème de la Variété des sources de données et la Variété des formats de stockage cible de l' \mathcal{ED} . Dans cet article, nous présentons une nouvelle approche de conception d' \mathcal{ED} guidée par la valeur et la variété, où différents scénarii d'intégration des \mathcal{LOD} impactant les processus ETL sont étudiés. Finalement, nous exposons des expérimentations validant nos propositions.

1 Introduction

La compétitivité dans le développement des pays, des régions, des universités et des entreprises est devenue un défi majeur à l'heure de l'intégration et de la mondialisation des économies en croissance. Le besoin à l'heure actuelle pour chaque entreprise est lié à la productivité de la chaîne de valeur ajoutée. Une des dimensions porteuses de valeur dans les entreprises et concentrant d'importants investissements est liée aux technologies de traitement des données. Les entrepôts de données (\mathcal{ED}) et leur environnement BI ont été créés pour répondre à un besoin de création de valeur ajoutée pour l'entreprise, à travers diverses possibilités analytiques. Depuis l'apparition des données massives (Big Data) avec ses principaux Vs (Volume, Variété, Vitesse, Visibilité, Véracité et Valeur), une baisse d'intérêt pour les \mathcal{ED} a été ressentie, de la même manière que les SGBD relationnels après l'apparition de NoSQL. Le cours des événements nous a appris que les SGBD relationnels continuent d'occuper une place importante et peuvent même cohabiter avec les SGBD NoSQL. Le challenge actuel des \mathcal{ED} est de se positionner dans ce nouvel environnement du Big Data, en tirant profit de ses Vs, pour susciter un regain d'intérêt qui contribuera à leur relance.

L'arrivée de l'ère *Linked Open Data* (\mathcal{LOD}) représente une excellente opportunité, car elles sont porteuses de valeur ajoutée, complémentaire à celle extraite à partir des sources de