

Mémoire Master 2 MIAGE classique

Thème

Etude des approches de modélisation des entrepôts de données (DataWarehouse)

Réalisé par :

Zakaria MEDJIR

Encadré par :

Monsieur Reda BENDRAOU

Promotion : 2017-2018

Introduction

Aujourd'hui, parmi les défis auxquels font face les entreprises, on trouve l'exploitation et l'analyse de données opérationnelles qu'elles détiennent dans leurs sources de données hétérogènes. Le but ultime de ces tâches est d'obtenir de l'information utile pour la prise de décision.

La Business intelligence ou système décisionnel est l'anneau manquant qui peut transformer ces données brutes en informations utiles et pertinentes qui peuvent supporter les décisions prises par les dirigeants des entreprises. Un concept central dans un tel système est l'entrepôt de données (Datawarehouse). Ce dernier est donc un composant principal du système décisionnel qui a pour but de stocker les données opérationnelles, provenant de plusieurs sources, dans une perspective décisionnelle et de les fournir aux utilisateurs sous certaines formes pour des fins d'analyse.

La mise en place d'un entrepôt de données nécessite une approche de modélisation qui prend en considération tous les aspects de développement comme la modélisation de données, la gestion de projet, la gestion des risques, le déploiement et bien d'autres aspects essentiels. Depuis des années, deux approches s'affrontent quant à la modélisation des entrepôts de données : l'approche de modélisation par sujet d'Inmon et l'approche dimensionnelle de Kimball. Cependant, ces dernières années, une troisième approche est apparue et elle gagne du terrain d'année en année. Cette approche est créée par Linstedt et s'appelle « Data Vault ». Dans le cadre de ce mémoire de Master, nous élaborons une étude comparative de ces trois approches de modélisation d'entrepôt de données.

La mise en place d'un entrepôt de données nécessite une approche de modélisation qui prend en considération tous les aspects de développement comme la modélisation de données, la gestion de projet, la gestion des risques, le déploiement et bien d'autres aspects essentiels.

Depuis des années, deux approches s'affrontent quant à la modélisation des entrepôts de données : l'approche de modélisation par sujet **d'Inmon** et l'approche dimensionnelle de **Kimball**. Cependant, une troisième approche est apparue et elle gagne du terrain d'année en année. Cette approche est créée par Linstedt et s'appelle « **Data Vault** »

Le but de ce travail est de faire une synthèse sur la BI à travers ces trois approches. D'abord, je ferai une étude approfondie sur les systèmes décisionnels et l'architecture autour de laquelle ils sont construits, en détaillant chacun de ses composants et en mettant l'accent sur l'entrepôt de données. Ensuite je vais me focaliser sur la présentation de chaque approche : sa définition,

sa philosophie, son architecture ainsi que sa méthodologie de développement. Enfin, je dresserai une analyse comparative entre les trois approches en s'appuyant sur des critères tels que :

- La méthodologie et l'architecture : structure & architecture, méthodologie de développement , coût et temps de déploiement... etc
- La modélisation de données : orientation des données, les outils utilisés, l'implication de l'utilisateur final ... etc
- Intégration de données : intégration des sources multiples, complexité de processus ETL ... etc
- Management du cycle du vie : facilité du changement du modèle, performances d'interrogation ... etc

Cette étude comparative sera accompagnée par des tests sur des données réelles (au sein de l'entreprise de mon contrat professionnalisation) avec un groupe témoin afin d'aboutir à une synthèse permettant de répondre à la question : « **Quelle approche dans quelle situation ?** ».

Bibliographie

[Weir2008] Weir : A Configuration Approach for Selecting a Data Warehouse Architecture, Thesis, 2008

(Kimball et Ross, 2013) Kimball, Ralph ; Ross, Margy : The data warehouse toolkit : The definitive guide to dimensional modeling. John Wiley & Sons, 2013.

[Denis 2008] Denis : Conception et réalisation d'un entrepôt de données institutionnel dans une perspective de support à la prise de décision, Thesis, 2008.

[Teste2009] Teste, Olivier : Modélisation et manipulation des systèmes OLAP : de l'intégration des documents à l'utilisateur, Université Paul Sabatier-Toulouse III, Dissertation, 2009

[Adamson2012] Adamson, Christopher : Mastering data warehouse aggregates : solutions for star schema performance. John Wiley Sons, 2012.

[Awel2014] Awel : Data Vault Modelling, Thesis,2014

[Awel2014] Mathiews, Diane : Data vault et bi. URL :<http://fr.slideshare.net/dlinstedt/prsentation-data-vault-et-bi-v20120508>

[Orlov2014] Orlov, Vadim : Data Warehouse Architecture : InmonCIF, Kimball Dimensional or Linstedt DataVault ? 2014. –URL : <https://blog.westmonroepartners.com/data-warehouse-architecture-inmon-cif-kimball-dimensional-or-linstedt-data-vault/>

(Poletto, 2012) Poletto, Maxime : L’informatique décisionnelle, These professionnelle, 2012. – URL <http://news.exia.cesi.fr/wp-content/uploads/2012/06/Maxime-Poletto-Th%C3%A8se.pdf>.