

# SYNTHÈSE BIBLIOGRAPHIQUE

Nettoyage des données en présence  
de données de référence

# INTRODUCTION

- ▶ Manque de qualité de données coûte en environ 600 milliards de dollars
- ▶ Data Cleaning
- ▶ Enjeu de la qualité des données

# DÉMARCHE

- ▶ Progressive et critique
- ▶ Arguments technologiques VS Arguments commerciaux

# LE MASTER DATA MANAGEMENT

- ▶ Objectif : définir des concepts et méthodes visant à établir au sein d'un SI un schéma de BDD de référence fiable.
- ▶ Répond à la problématique de qualité de données : limite l'entropie des données
- ▶ Pas de nettoyage des données

# LES DONNÉES

- ▶ Données transactionnelles
- ▶ Données analytiques
- ▶ Données de référence
- ▶ Définition d'un parc de données de référence de qualité

# FONCTIONNEMENT AU SEIN D'UN SI D'ENTREPRISE

- ▶ BDD unique et unifiée pour l'ensemble des applications opérationnelles du S.I.
- ▶ Maintenance des données simplifiées
- ▶ Conditions nécessaires au bon fonctionnement :
  - ▶ Stockage des données unique
  - ▶ Système maître des données
  - ▶ Les données du système de M.D.M sont des données de références

# FONCTIONNEMENT AU SEIN D'UN SI D'ENTREPRISE

## ► Modules d'un système de M.D.M. :

Gestion du cycle de vie	Administration
Stockage	Gestion des méta-données
Gestion des accès aux données	Règles directives

## ► Processus d'un système de M.D.M. :

Management	Consolider
Gouverner et nettoyer	Partager
Exploiter	

# GESTION ET IMPACT SELON LE TYPE DE DONNÉES

- ▶ TIERS : données liées à des personnes physiques, morales
  - ▶ Principale difficulté : la consolidation
- ▶ PRODUITS :
  - ▶ Principale difficulté : gestion collaborative
- ▶ FINANCIÈRES : établissement des comptes annuels, rapport de gestion
  - ▶ Utilisation de modèles hiérarchiques



# POSITIONNEMENT AU SEIN DU SI D'ENTREPRISE

- ▶ ” La seule source d'erreur possible dans un ordinateur se trouve entre la chaise et le clavier !”.
- ▶ Problématique d'intégration n-carrée
- ▶ Corruption lors de la propagation des données
- ▶ Maintenance difficile et qualité des données faibles
- ▶ Contrôler et uniformiser les modes de communication entre les BDDs

# E.A.I.

- ▶ Applications Opérationnelles
- ▶ Synchronisation des données
- ▶ Fonctionnalités des applications métiers
- ▶ Limites

# LES OFFRES DU MARCHÉ

► ORACLE

► IBM

# LES CHALLENGES

- ▶ Intégration au sein des processus d'entreprise
- ▶ Nettoyage des données

# CONCLUSION

- ▶ Gestion de la qualité : problème d'actualité
- ▶ Arguments commerciaux
- ▶ Le M.D.M. concept relativement nouveau
- ▶ Défis liés à sa complexité

# RÉFÉRENCES

- ▶ BRASSEUR Christophe. Data Management : qualité des données et compétitivité. Paris : Hermès, 2005.
- ▶ Hong Hai Do Erhard Rahm. Data cleaning : Problems and current approaches.  
[http://wwwiti.cs.uni-magdeburg.de/iti\\_db/lehre/dw/paper/data\\_cleaning.pdf](http://wwwiti.cs.uni-magdeburg.de/iti_db/lehre/dw/paper/data_cleaning.pdf).
- ▶ Wenfei Fa et al. Interaction between record matching and data repairing, 2011.
- ▶ Wenfei Fan. Dependencies revisited for improving data quality.  
<http://www.lifl.fr/~bonifati/teaching/dq/lucidi/pods08Tutorial.pdf>.
- ▶ IBM. An IBM White Paper : How Master Data Management Serves the Business, Novembre 2011.
- ▶ Divesh Srivastava Lukasz Golab, Flip Korn. Discovering pattern tableaux for data quality analysis : a case study.  
<http://qdb2011.dia.uniroma3.it/participants/program/p47-GOLAB.PDF>.
- ▶ ORACLE. An Oracle White Paper : Master Data Management, Septembre 2011.