



MySQL Workbench

Guide de modélisation des données pour les développeurs et les DBA

Livre blanc MySQL® pour l'entreprise

Table des matières

1	INTRODUCTION	3
2	TYPES DE DONNÉES À GÉRER	3
3	GESTION DES DONNÉES BASÉE SUR DES MODÈLES	4
4	AVANTAGES RELATIFS À LA GESTION DES DONNÉES BASÉE SUR DES MODÈLES...	5
4.1	Gestion des métadonnées	6
4.2	Livraison rapide d'applications	6
4.3	Gestion des changements	7
4.4	Gestion des applications intégrées	7
4.5	Rapports et communication	8
4.6	Gestion de la performance.....	8
5	MYSQL WORKBENCH – MODÉLISATION DES DONNÉES/CONCEPTION DE MYSQL	9
5.1	Conception de bases de données	9
5.2	Pro-ingénierie et rétro-ingénierie	11
5.3	Gestion des changements	12
5.4	Rapports et documentation.....	13
5.5	Fonctionnalités de MySQL Workbench.....	14
6	CONCLUSION.....	14

1 Introduction

Les données sont l'élément moteur de toutes les entreprises prospères, quelle que soit leur taille et leur secteur d'activité. Les données d'une entreprise revêtent plusieurs formes : des nouvelles données transactionnelles sous forme de commandes entrantes à partir d'un site Internet, des données d'informatique décisionnelle glanées à partir des données des clients et aidant le personnel d'encadrement d'une société à prendre des décisions stratégiques intelligentes, des informations d'historique indispensables aux responsables de la conformité ou encore des métadonnées qui décrivent les différents éléments de données composant les systèmes pilotés par des données ainsi que l'utilisation de ces éléments. Si une entreprise traite les données clés de façon incorrecte ou si elle les perd, cette expérience peut souvent être désastreuse et engendrer d'importantes pertes financières ainsi qu'une dégradation de sa réputation.

Dès lors, les entreprises modernes et prospères ne laissent rien au hasard en termes de définition, de conception et de mise en œuvre de leurs données. Cela est comparable à une approche professionnelle et basée sur les processus en matière de création et de gestion des données qui circulent dans ses systèmes d'entreprise – un processus géré par des experts de la gestion de données, qui utilisent une approche des données basée sur les modèles ainsi que des outils adaptés dans le processus afin de garantir que la saisie et l'administration des données sont correctement effectuées.

Ce livre blanc décrit les différents types de données dont les entreprises ont besoin pour sa gestion. Il étudie également les raisons légitimant la nécessité d'une approche de la gestion des données basée sur des modèles, et il expose les avantages d'une telle approche. De plus, il présente la façon dont le module de modélisation des données du produit MySQL Workbench peut être une aide indispensable dans les mains expertes des modélisateurs de données, des développeurs et des DBA qui sont chargés de gérer l'infrastructure de gestion des données complexe d'une entreprise dynamique et en pleine croissance.

2 Types de données à gérer

Bien qu'il existe d'innombrables façons de classer les données, la plupart des formes de données que les entreprises modernes gèrent sont réparties dans sept groupes :

1. **Données opérationnelles** : normalement, les données de traitement transactionnel existant sous la forme de commandes client nouvelles/mises à jour et les autres données qui prennent en charge les produits et services vendus par les entreprises. Ces données sont généralement présentes dans les bases de données relationnelles prenant en charge les flux de données transactionnelles.
2. **Données d'informatique décisionnelle** : elles existent sous forme de données opérationnelles passées et présentes, qui sont utilisées pour comprendre certains aspects, comme les tendances d'achat des clients, l'impact des programmes marketing, etc. Ces données résident généralement dans des zones de préproduction, connues sous le nom d'entrepôts de données ou de magasins de données analytiques, et elles sont séparées des données opérationnelles afin d'améliorer les temps de réponse de ces systèmes.
3. **Données historiques** : elles représentent l'activité d'historique des systèmes d'entreprise ou des journaux d'audit de l'utilisation des données au sein d'une organisation. Elles diffèrent des données d'informatique décisionnelle dans le sens où l'on y accède rarement et elles sont principalement conservées en ligne afin de respecter les réglementations en matière de conformité du gouvernement et de l'industrie.
4. **Données d'intégration** : elles sont utilisées pour gérer le flux de données des systèmes opérationnels vers les magasins de données analytiques ou historiques. Elles définissent le plus souvent le processus de transformation des données transactionnelles en données d'informatique décisionnelle.
5. **Données maîtres** : elles sont également appelées « données de référence ». Les données de référence ne dépendent pas des autres éléments de données pour leur identité ou leur signification.

et elles sont souvent utilisées comme données de consensus partagées de façon homogène sur l'ensemble des systèmes.

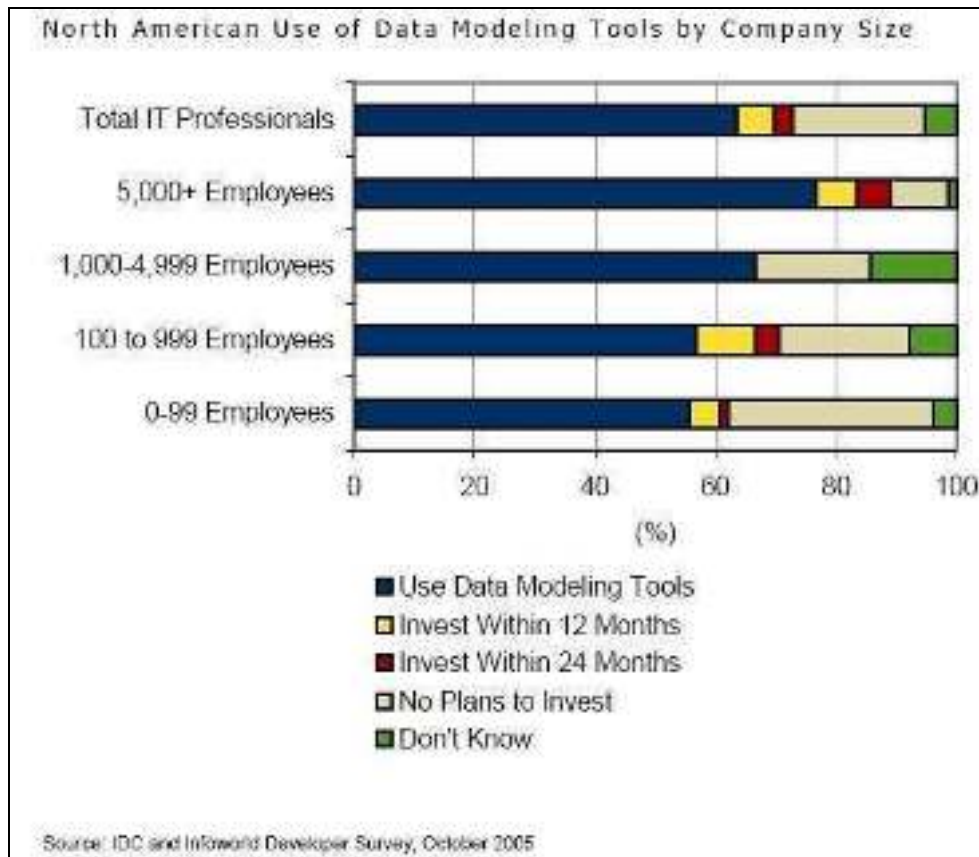
6. **Métadonnées** : ce sont des « données sur les données » qui servent de point de définition des éléments de données ainsi que comme description de la façon dont elles doivent être utilisées.
7. **Données non structurées** : elles sont généralement gérées dans les systèmes de gestion de contenu (bien qu'elles soient parfois déplacées vers les moteurs SGBDR traditionnels), administrant l'évolution du cycle de vie des informations numériques (documents, fichiers vidéo, etc.)

Une définition plus rigoureuse des classements de données peut exister, mais les descriptions ci-dessus représentent ce à quoi l'entreprise moderne doit faire face actuellement dans le domaine de la gestion des données.

3 Gestion des données basée sur des modèles

L'approche reconnue par les experts en données comme étant la meilleure méthode en matière de gestion de la saisie, de la définition et de la mise en œuvre des données au sein d'une organisation est la méthode basée sur des modèles. C'est-à-dire qu'elle dépend de la modélisation de l'utilisation et des relations des données qui existent dans les catégories de données abordées dans la section précédente. Cela s'applique aux systèmes existants, mais aussi aux systèmes qui sont en cours de construction.

Les modèles sont le meilleur moyen de représenter la définition des éléments de données qui prennent en charge les différents magasins de données présents au sein d'une entreprise. Par conséquent, il n'est pas étonnant que la plupart des sociétés d'informatique utilisent des pratiques comme le schéma du modèle entité-association (ERD, entity relationship diagramming) ou d'autres formes de modélisation permettant de saisir et de protéger leurs structures de données. La pratique de la gestion des données basée sur des modèles a été présentée fin 2005 dans une étude de développeur nord-américaine, menée conjointement par IDC et Infoworld, montrant que la plupart des entreprises font appel à des outils de modélisation pour pouvoir saisir et mettre en œuvre les données :



Il est intéressant de noter que même les entreprises de petite taille (en termes du nombre d'employés) utilisent une approche basée sur des modèles afin de gérer leurs données. Notons également que près de 3/4 des grandes entreprises utilisent actuellement une approche basée sur des modèles ou qu'elles sont en passe de le faire.

Une utilisation aussi importante sous-entend que l'utilisation d'une approche de la gestion des données basée sur des modèles présente des avantages réels et tangibles. Il peut être utile de décrire brièvement ces avantages, puis d'approfondir chacun d'entre eux pour comprendre pourquoi une approche de la gestion des données basée sur des modèles est préférée à une autre approche.

4 Avantages relatifs à la gestion des données basée sur des modèles

La gestion des données basée sur des modèles apporte de nombreux avantages, dont les plus concrets sont les suivants :

- **Gestion des métadonnées** : elle garantit la cohérence des données, applique les normes des éléments de données utilisés au sein d'une organisation et permet d'identifier et de classer les éléments de la gouvernance des données.
- **Livraison rapide d'applications** : réduction du délai de réalisation et de mise en œuvre d'une nouvelle conception de données physiques, ainsi que de l'application qui utilise la base de données sous-jacente.
- **Gestion des changements** : elle permet de gérer le changement entre les différentes itérations des conceptions de données.

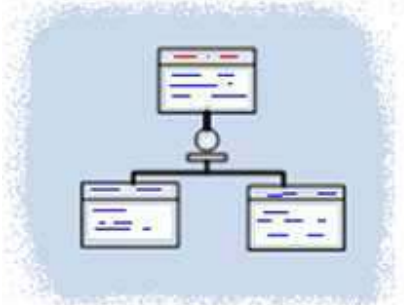
- **Gestion des applications intégrées** : elle élimine l'aspect 'boîte noire des applications intégrées en offrant un rendu graphique au cœur de chaque application, qui est la base de données.
- **Rapports et communication** : grande simplification de la communication et des rapports sur les conceptions de données nouvelles et modifiées.
- **Gestion de la performance** : elle permet d'identifier plus rapidement les défauts de création des conceptions de données, à l'origine de l'inefficacité des temps de réponse dans les systèmes pilotés par des données.

Chacun de ces domaines sera exploré plus en détail dans les sections suivantes.

4.1 Gestion des métadonnées

L'importance de disposer de définitions d'éléments de données correctement établies et normalisées a été parfaitement saisie par les entreprises qui se sont débattues avec la difficulté de tracer et de préparer l'utilisation des artefacts de données signifiant et se rapportant à la même chose, mais qui sont définis de façon différente sur plusieurs applications et systèmes. Ainsi, garantir qu'un élément de données appelé CUSTOMER_ID est défini de façon cohérente sur tous les systèmes qui l'utilisent (par exemple le type de données correct, la signification, etc.) atténue la peine de devoir l'utiliser dans les applications futures ou dans les magasins de données analytiques qui sont alimentés par de nombreux systèmes opérationnels.

De plus, les réglementations en matière de conformité du gouvernement et de l'industrie ont créé un nouvel ensemble de raisons justifiant le caractère fondamental d'une gestion de données correcte. Les besoins actuels des entreprises modernes sont les suivants : la nécessité d'assurer des données d'une qualité la plus élevée possible (aucune donnée non valide, correspondance entre les données et leur objectif et 100 % de définition), la confidentialité des données (internes uniquement ou externes), la sécurité des données (hautement sensibles, non sensibles, etc.), le secteur d'activité qui possède l'élément de données (par exemple, finance, etc.) et la garantie de la conformité (Sarbanes Oxley, etc.).



Une approche de la gestion des métadonnées basée sur des modèles permet aux architectes de données et aux autres personnes chargés de la gestion des données d'une organisation de saisir et de classer tous les aspects des éléments de données. Les modèles et les outils qui simplifient la gestion des métadonnées permettent également de référencer, d'effectuer des rapports et des recherches, ainsi que d'examiner les aspects du contrôle des changements de tous les éléments de données qui y sont définis.

4.2 Livraison rapide d'applications

Le cœur de la grande majorité des applications modernes, c'est la base de données sous-jacente. Par conséquent, il est essentiel qu'aucun retard n'apparaisse lors de la conception et de la mise en œuvre physique des structures de données qui prendront en charge les applications d'entreprise critiques.

Aucun substitut n'existe à ce qu'une approche basée sur des modèles offre lorsqu'il s'agit de livrer rapidement des conceptions de bases de données physiques de qualité. Souvent, une conception de base de données commence de manière conceptuelle, c'est-à-dire qu'il s'agit de définir les entités de données, leurs relations et les attributs de façon à ne pas cibler la plateforme d'un fournisseur de bases de données physiques. Une fois qu'un modèle de données logique a été établi, il est alors transformé en une conception physique qui est dirigée vers un SGBDR spécifique (par exemple, MySQL, Oracle, etc.). Le modèle logique est souvent modifié afin de répondre aux exigences d'une plateforme physique particulière et à des fins de performance (ajout d'un partitionnement aux tables, etc.)

Ceci est un exemple, cliquez sur le lien de téléchargement pour obtenir le cours complet.

