



Activité 2

MAÎTRISER LES DIMENSIONS

Compétences visées :

- Reprendre les faits
- Comprendre les dimensions
- Appréhender la dimension date
- Comprendre les dimensions conformes
- Comprendre les dimensions dégénérées
- Comprendre les notions de Junk dimension et Role-playing dimension

Recommandations clés :

- Répondre à chaque question directe
- Argumenter la réponse
- Revoir la partie du cours concernée par la question
- Définir chaque mot clé de la question ainsi que la réponse
- Identifier les dimensions
- Utiliser MySQL dans la création des tables de faits



Exercice 1

1. Pourquoi utilisons-nous la dimension dans notre modèle de données ?

- ☐ Il aide dans les opérations de mise à jour et d'insertion
- ☒ Il contribue à la performance et à la convivialité
- ☐ Il aide à traiter des données hétérogènes

Argumenter et justifier la réponse depuis la présentation PPT ?

La dimension dans un modèle de données est souvent utilisée pour améliorer la performance en permettant une organisation logique et structurée des données. Cela facilite également leur manipulation et leur compréhension par les utilisateurs, ce qui contribue à la convivialité du système.

Exercice 1

2. Quel est le format de clé courant pour les dimensions de date ?
- ☐ Nous utilisons couramment une clé de substitution standard comme 1, 2, 3, 4,...
 - ☒ Nous utilisons généralement une clé de substitution significative au format AAAAMMJJ.
 - ☐ Nous n'utilisons généralement pas de clé de substitution, mais uniquement la colonne de date au format "AAAA-MM-JJ".

Argumenter et justifier la réponse depuis la présentation PPT ?

Dans de nombreux systèmes, une clé de substitution significative est utilisée pour les dimensions de date. Cette clé est généralement au format AAAAMMJJ (année, mois, jour) et permet une représentation facile à comprendre et à manipuler des dates.

Exercice 1

3. Dans la table de faits de traitement des commandes, nous avons plusieurs horodatages pour différentes étapes du traitement des commandes. Les analystes/utilisateurs métier veulent analyser les données en fonction de tous ces différents horodatages. Comment pouvons- nous répondre à leur exigence?

- ☒ Utiliser une dimension de jeu de rôle
- ☐ Utilisation d'une dimension dégénérée
- ☐ Utilisation d'une dimension indésirable
- ☐ Utilisation d'une dimension conforme

Argumenter et justifier la réponse depuis la présentation PPT ?

Une dimension de jeu de rôle est utilisée pour représenter différentes interprétations ou perspectives sur les mêmes données de faits. Dans ce scénario, où différents horodatages pour différentes étapes du traitement des commandes sont nécessaires pour l'analyse, une dimension de jeu de rôle pourrait être utilisée pour représenter ces différents horodatages, permettant ainsi aux analystes/utilisateurs métier d'analyser les données en fonction de leurs besoins spécifiques.

Exercice 1

4. Qu'est-ce qu'une dimension conforme ?

- ☒ Une dimension qui est partagée par plusieurs tables de faits.
- ☐ Une dimension qui contient des attributs textuels.
- ☐ Une dimension qui contient des attributs numériques.
- ☐ Une dimension qui est liée à une seule table de faits.

Argumenter et justifier la réponse depuis la présentation PPT ?

Une dimension conforme est une dimension qui peut être utilisée avec plusieurs tables de faits. Cela signifie que les clés étrangères de différentes tables de faits peuvent être reliées à cette dimension. Cela permet une utilisation cohérente et efficace des dimensions communes dans plusieurs contextes d'analyse de données.

Exercice 1

5. Qu'est-ce qu'une dimension dégénérée ?

- ☐ Une dimension qui contient des valeurs nulles.
- ☐ Une dimension qui est associée à des tables de référence.
- ☐ Une dimension qui est stockée en tant que clé étrangère dans la table de faits.
- ☒ Une dimension qui est représentée par un attribut de la table de faits.

Argumenter et justifier la réponse depuis la présentation PPT ?

Une dimension dégénérée est une dimension qui est représentée directement par un attribut (ou plusieurs attributs) de la table de faits elle-même, au lieu d'être liée à une table de dimension distincte. Cela se produit généralement lorsque les attributs de la dimension sont nécessaires pour l'analyse mais qu'ils ne justifient pas la création d'une table de dimension séparée.

Exercice 1

6. Qu'est-ce qu'une junk dimension ?

- ☐ Une dimension qui contient des données redondantes.
- ☒ Une dimension qui contient des attributs inutiles.
- ☐ Une dimension qui est utilisée pour stocker des valeurs nulles.
- ☐ Une dimension qui est utilisée pour représenter des données de faible granularité.

Argumenter et justifier la réponse depuis la présentation PPT ?

Une junk dimension (ou dimension poubelle) est une technique de modélisation de données dans laquelle des attributs qui ne sont pas liés les uns aux autres sont regroupés dans une seule table de dimension. Ces attributs peuvent être des indicateurs binaires ou des attributs catégoriels qui ne s'intègrent pas facilement dans d'autres dimensions. Cette technique permet de réduire la complexité du modèle de données en regroupant ces attributs moins importants dans une seule table, ce qui peut améliorer la performance des requêtes.

Exercice 1

7. Qu'est-ce qu'une role-playing dimension ?

- ☐ Une dimension qui contient des rôles d'utilisateurs.
- ☐ Une dimension qui est utilisée pour stocker des valeurs numériques.
- ☐ Une dimension qui est utilisée dans plusieurs modèles dimensionnels.
- ☒ Une dimension qui est utilisée pour représenter différentes perspectives d'une même entité

Argumenter et justifier la réponse depuis la présentation PPT ?

Une dimension de jeu de rôle (role-playing dimension) est une dimension utilisée dans un modèle de données pour représenter différentes perspectives ou "rôles" d'une même entité. Par exemple, une dimension de temps pourrait être utilisée dans un entrepôt de données pour représenter à la fois la date de commande et la date de livraison dans une table de faits des commandes. Chaque utilisation de la dimension de temps dans ce contexte représente un "rôle" différent de cette dimension. Cela permet aux analystes d'interroger les données selon différentes perspectives sans avoir à créer des dimensions distinctes pour chaque rôle.

Exercice 1

8. Un Data Warehouse contient deux tables de faits différentes, "Ventes" et "Commandes". Quelle approche de modélisation dimensionnelle serait la plus appropriée pour partager une dimension "Produit" entre ces deux tables de faits ?

- ☒ Dimension conforme
- ☐ Dimension dégénérée
- ☐ Junk dimension
- ☐ Role-playing dimension

Argumenter et justifier la réponse depuis la présentation PPT ?

Une dimension conforme est utilisée lorsqu'une dimension est partagée par plusieurs tables de faits. Dans ce cas, la dimension "Produit" pourrait être liée aux tables de faits "Ventes" et "Commandes" grâce à des clés étrangères. Cela permet d'assurer la cohérence des données et d'éviter la redondance dans le modèle de données.

Exercice 1

9. Vous avez un ensemble de données comportant des attributs de date tels que "Date de commande", "Date de livraison" et "Date de paiement". Comment modéliseriez-vous ces attributs de date dans un modèle dimensionnel ?
- ☐ En créant une dimension de date unique qui est partagée par toutes les tables de faits.
 - ☒ En créant des dimensions de date séparées pour chaque attribut de date.
 - ☐ En utilisant une dimension dégénérée pour chaque attribut de date.
 - ☐ En utilisant une junk dimension pour regrouper tous les attributs de date.

Argumenter et justifier la réponse depuis la présentation PPT ?

Chaque attribut de date, tel que "Date de commande", "Date de livraison" et "Date de paiement", devrait être associé à sa propre dimension de date distincte. Cela permettra aux utilisateurs d'analyser les données en fonction de différents horodatages de manière indépendante.

Exercice 1

10. Vous souhaitez suivre les performances des employés dans différentes régions géographiques. Chaque employé peut être affecté à plusieurs régions. Quel type de dimension serait le plus approprié pour représenter cette relation ?

- ☐ Dimension conforme
- ☐ Dimension dégénérée
- ☒ Junk dimension
- ☐ Role-playing dimension

Argumenter et justifier la réponse depuis la présentation PPT ?

Une dimension poubelle (Junk dimension) est utilisée pour regrouper des attributs non corrélés dans une seule dimension. Dans ce cas, la dimension poubelle pourrait contenir les associations entre les employés et les régions géographiques. Cela permet de gérer efficacement les relations multiples entre les employés et les régions sans avoir à créer une dimension complexe ou à introduire des redondances dans le modèle de données.

Exercice 2

On va travailler sur la même étude de cas « E-commerce », mais on va se focaliser sur les dimensions.

Comme nous avons vu dans le cours, la dimension date est généralement la dimension la plus importante.

Date_PK	Date	Jour_ semaine	Jour_semaine_ numéro	Jour_ mois	Jour_ année	Semaine_ année	Mois_ num	Mois_ nom	trimestre	Année
20100101	2010-01-01	vendredi	5	1	1	53	1	Janvier	1	2010
...

Exercice 2 : Suite

1. Créez la table `date_dim` sans oublier les contraintes (aucun attribut ne doit être null).
2. Ajoutez la contrainte de la clé primaire.
3. Afin d'optimiser la performance des requêtes, ajoutez un INDEX à la colonne `Date`.
4. On veut Générer des dates d'une année à partir du 2024-01-01 jusqu'à 2024-12-31 et les insérer dans la table `date_dim`, pour faire cela suivons les étapes suivantes:
 1. Créez une table 'nombres' avec une seule colonne 'id'
 2. Créer une procédure `genererNombre(start idDepart, idFin)` qui va insérer les nombres de 0 jusqu'à 400 dans la table `nombres`
 3. Appeler la procédure en lui passant 0 et 365 comme paramètres
 4. Vérifier que les enregistrements ont été bien insérés
 5. Exécuter la requête `SELECT n, DATE_ADD('2024-01-01', INTERVAL n DAY) AS dateGeneree FROM numbers;`
 6. A quoi sert la fonction `DATE_ADD` ?
 7. Adapter la requête pour retenir seulement les dates inférieures ou égales à 2024-12-31

Exercice 2 : Suite

4. On veut Générer des dates d'une année à partir du 2024-01-01 jusqu'à 2024-12-31 et les insérer dans la table date_dim, pour faire cela suivons les étapes suivantes:
 7. Rappelez vous de l'une des syntaxes des requêtes imbriquées :
 1. `select columnsNames from (select * from tableName) as resultatSousRequete`
 8. On veut maintenant réaliser une requete qui affichera les colonnes n, dateGeneree et le nom du jour de la semaine,
 1. Quelle fonction permet d'avoir le nom du jour d'une date ?
 2. créer et tester la requete.
 9. Adapter la requête pour ajouter les autres colonnes, tester et afficher la requete
 10. C'est le temps pour faire la combinaison de requete insert et select pour inserer le resultat De select dans la table date_dim
 11. On veut afficher une colonne trimestre, et une colonne isWeekEnd pour savoir respectivement le trimestre d'une date et s'elle correspond à un jour de fin de semaine (samedi et dimanche), donne une requete select qui répond à la question.

	n	generated_date	dayName
►	0	2024-01-01	Monday
	1	2024-01-02	Tuesday
	2	2024-01-03	Wednesday
	3	2024-01-04	Thursday
	4	2024-01-05	Friday

Exercice 3 :

Considérons un système de suivi des performances sportives. Les données disponibles sont les suivantes :

- ID de match
- Date du match
- Nom de l'équipe à domicile
- Nom de l'équipe à l'extérieur
- Score de l'équipe à domicile
- Score de l'équipe à l'extérieur
- ID du joueur
- Nom du joueur
- Position du joueur
- Temps de jeu du joueur dans le match

Exercice 3 : Suite

1. Identifiez les faits et les dimensions.
2. Proposez une modélisation dimensionnelle du problème.
3. Créez la table de faits "Performances" avec les colonnes nécessaires pour stocker les données en utilisant SQL.
4. Identifiez les colonnes qui serviront de clés étrangères pour les dimensions.
5. Proposez les définitions des tables de dimensions associées en utilisant SQL.
6. Ajoutez des contraintes de clé étrangère pour les colonnes correspondantes dans la table de faits en utilisant SQL.