

Série n° 1 en Entrepôts de données : Modélisation Multidimensionnelle

Exercice 1 (Modélisation de base)

On veut analyser la performance de coureurs qui se préparent à des compétitions sportives à travers les entraînements. Un entraînement consiste à effectuer plusieurs tours d'un terrain. Un coureur est performant s'il fait beaucoup de tours de terrain et s'il réalise un temps moyen minimal par tour. Les entraînements se font dans plusieurs terrains de plusieurs villes où chaque terrain appartient à une seule ville. Chaque terrain est caractérisé par une adresse et par une superficie. Chaque coureur possède un nom, un prénom et une date de naissance et il appartient à un groupe possédant une date de création. Aussi, chaque coureur peut être classé dans l'une de deux catégories désignées par «A» si l'âge est compris entre 18 et 40 ans et par «B» si l'âge est supérieur à 40 ans. Enfin, on veut analyser les entraînements par jour, par mois et par année.

1. Représenter le schéma en étoile.
2. Analyser l'additivité de la (des) mesure(s) identifiées.
3. Transformer le schéma précédent en schéma en flocon de neiges.

Exercice 2 (Hiérarchies et conformité)

On veut analyser les projections de films dans les salles de cinéma. L'entrepôt de données des projections est alimenté à travers une page Web par des spectateurs qui ont suivi les films. L'analyse doit permettre de connaître les projections les mieux notées, les périodes de temps qui captent plus de spectateurs, les meilleurs films et les salles les plus fréquentées. Pour cela, chaque spectateur renseigne la durée de suivi d'un film et il attribue un nombre d'étoiles à la projection (pas d'étoiles : mauvaise projection, une étoile : projection normale, deux étoiles : bonne projection, trois étoiles : excellente projection). Chaque spectateur peut regarder le même film dans des salles différentes et à des dates différentes. Cependant, il ne peut renseigner les informations d'un film projeté à une date donnée dans une salle donnée qu'une seule fois. Chaque spectateur est identifié par une adresse e-mail et possède un nom et un prénom. Chaque salle possède un nom, une adresse et une capacité et est située dans une ville, décrite par un nom et un nombre d'habitants. L'analyse concerne plusieurs villes d'un seul pays et s'effectue par jour, par mois, par année. Enfin, chaque film possède une durée, un titre et peut être classé par genre (action, horreur, ...) et pour chaque genre de film, on note l'âge minimum de visualisation.

1. Représenter le schéma en flocon de neige de ce cas.
2. Analyser l'additivité de la (des) mesure(s) identifiées.
3. Supposons qu'on veut analyser les projections selon la nature de la journée aussi (journée de travail, jour férié ou week-end). Modifier le schéma pour répondre à ce changement.
4. Supposons que maintenant, on veut connaître les trimestres, les semestres et les semaines qui captent plus de spectateurs. Modifier le schéma pour tenir compte de ce besoin.
5. Identifier les types de hiérarchies contenues dans le schéma en flocon de neige résultant.
6. On suppose que maintenant, les salles de cinéma appartiennent à des cités, qui à leur tour appartiennent aux villes mais que certaines salles restent affectées aux villes ;
 - a. Illustrer par un exemple l'appartenance des salles à des cités et l'appartenance de certains salles directement aux villes.
 - b. Proposer une solution pour un groupement correct des données.