

BD51

Business Intelligence & Data Warehouse

Examen Final A14

Rédacteur : Christian FISCHER

Conditions de réalisation :

Documents autorisés : Supports de Cours et TP de BD51

Durée : 2h

Barème : Le sujet comprend 2 dossiers indépendants.

Dossier 1	Requêtes SQL OLAP pour la SNCF	
12 pts	Rédiger les 8 requêtes SQL	12 pts
Dossier 2	Etude et Modélisation du Data Warehouse d'ESN	
8 pts	Modèle en étoile Gestion du calendrier Architecture logicielle	4 pts 2 pts 2 pts

Dossier 1 : Interrogation de l'entrepôt de données sur la fréquentation des trains en région parisienne

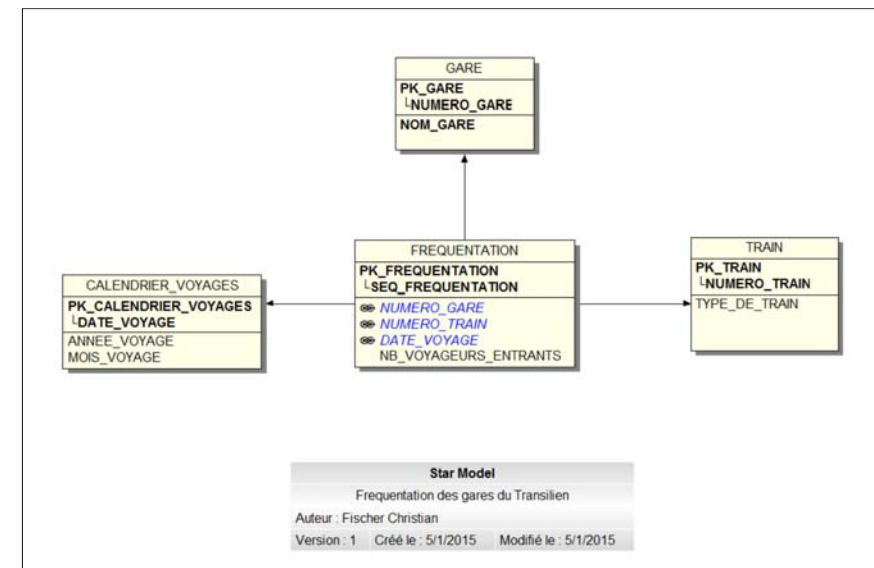
Présentation du contexte

Pour la SNCF, la valorisation des données est un levier d'innovation pour une mobilité plus informée, fluide et personnalisée. Déjà investi dans l'Open Data, SNCF propose aujourd'hui un challenge de data science dont l'objectif est de bâtir un modèle prédictif de la fréquentation des gares.

L'équipe chargée de l'entrepôt de données a récemment examiné la conception du système d'analyse actuel.

Mr Fischer C. a validé le modèle décisionnel décrivant le système d'information relatif à la fréquentation des gares.

Modèle en étoile validé



Ce modèle est géré par une base Oracle 11G Enterprise Edition avec l'option partitionnement de tables.

SEQ_FREQUENTATION = séquence Oracle pour identifier chaque fréquentation par jour, gare et train.

Travail à faire :

Rédiger les requêtes SQL ci-dessous.

Utilisation des jointures ANSI, des alias de tables et des alias de colonnes.

Requête 1 : (0,5 pt)

Calculer le nombre de voyageurs entrants par année pour 2014 et 2015

Requête 2 : (1 pt)

Calculer le nombre de voyageurs entrants par année et mois pour l'année 2014

Requête 3 : (1 pt)

Calculer le nombre de voyageurs entrants par année et mois pour l'année 2014 et 2015. Inclure le total annuel et le total général dans le résultat à l'aide de la clause Grouping Sets.

Requête 4 : (1,5 pt)

Calculer le nombre de voyageurs par gare pour l'année 2014.

Le résultat devra comporter le nom de la gare, le nombre de voyageurs et le classement de chaque gare en fonction de sa fréquentation

Exemple de résultat :

	Gare	Nb_entrants	Rang
1	LYCEE HENRI SELLIER	32116112474	1
2	LA REMISE A JORELLE	32116054413	2
3	NEMOURS SAINT-PIERRE	32007110740	3
4	MORET VENEUX LES SABLONS	32006436220	4
5	LE MEE	32006400085	5
6	PARIS GARE DE LYON	32006362802	6
7	VILLENEUVE TRIAGE	32006265035	7
8	MOULIN GALANT	32006116845	8
9	LE BRAS DE FER (EVRY GENOPOLE)	32006113925	9
10	SOUPPES CHATEAUN LONDON	32005220074	10
11	EVRY COURCOURONNES (CENTRE)	32004908546	11
12	BOIGNEVILLE	32004822878	12
13	VILLENEUVE SAINT-GEORGES	32004779861	13
14	BAGNEAUX SUR LOING	32004734470	14
15	MONTIGNY SUR LOING	32004705998	15

Requête 5 : (1,5 pt)

Calculer le nombre de voyageurs pour l'année 2014 et 2015 et par type de train.

Le résultat devra comporter l'année, le type de train, le nombre d'entrants et le classement en fonction de la fréquentation

Le résultat sera trié sur le nombre d'entrants.

Exemple de résultat :

	Année	TYPE_DE_TRAIN	nb_entrants	rang
1	2015	RER	2115982652996	1
2	2015	Train SNCF Banlieue	2081854303667	2
3	2015	Tramway	2081854303667	2
4	2015	TGV Duplex	2081854303667	2
5	2015	RER Rénové	2081854303667	2
6	2014	RER	2064373319844	6
7	2014	Tramway	2031077369289	7
8	2014	Train SNCF Banlieue	2031077369289	7
9	2014	RER Rénové	2031077369289	7
10	2014	TGV Duplex	2031077369289	7
11	2015	TGV	2013597591230	11
12	2014	TGV	1964485454721	12

Requête 6 : (1,5 pt)

Calculer le nombre de voyageurs pour l'année 2014 par type de train.

Le résultat devra comporter l'année, le type de train, le nombre d'entrants et les classements en fonction de la fréquentation (avec et sans égalité)

Le résultat sera trié sur le nombre d'entrants.

Exemple de résultat :

	TYPE_DE_TRAIN	nb_entrants	rang	drang
1	RER	2064373319844	1	1
2	RER Rénové	2031077369289	2	2
3	TGV Duplex	2031077369289	2	2
4	Train SNCF Banlieue	2031077369289	2	2
5	Tramway	2031077369289	2	2
6	TGV	1964485454721	6	3

Requête 7 : (2 pt)

Etape 1 :

Calculer le nombre de voyageurs pour l'année 2014 par mois.

Le résultat devra comporter le mois et le nombre d'entrants.

Etape 2 :

A partir de la requête précédente ajouter un cumul glissant sur 6 mois et une moyenne sur 6 mois.

Exemple de résultat :*Note : dans l'extrait les fréquentations sont identiques pour certains mois*

	mois	nb_entrants	cumul_6mois	moy_6mois
1	1	1032186659922	1032186659922	1032186659922
2	2	932298794799	1964485454721	982242727360
3	3	1032186659922	2996672114643	998890704881
4	4	998890709367	3995562824010	998890706002
5	5	1032186659922	5027749483932	1005549896786
6	6	998890709367	6026640193299	1004440032216
7	7	1032186659922	6026640193299	1004440032216
8	8	1032186659922	6126528058422	1021088009737
9	9	998890709367	6093232107867	1015538684644
10	10	1032186659922	6126528058422	1021088009737
11	11	998890709367	6093232107867	1015538684644
12	12	1032186659922	6126528058422	1021088009737

Requête 8 : (3 pt)

Etape 1 : Calculer le nombre de voyageurs par année de voyage et type de train pour les années 2014 et 2015

Seuls les trains de type TGV et RER doivent être inclus dans le résultat.

Le résultat est trié par Année et Type de train

Exemple de résultat :

	ANNEE_VOYA...	TYPE_DE_TR...	Nb_entrants
1	2014	RER	2064373319844
2	2014	TGV	1964485454721
3	2014	TGV Duplex	2031077369289
4	2015	RER	2115982652996
5	2015	TGV	2013597591230
6	2015	TGV Duplex	2081854303667

Etape 2 : Utilisation de l'opérateur PIVOT pour obtenir une seule ligne par année avec en colonnes le nombre de voyageurs par type de train.

Le total des voyageurs par type de train TGV et le total général doit être calculé.


Exemple de résultat :

	Année	TGV Simple	TGV Duplex	Total TGV	RER	Total
1	2014	1964485454721	2031077369289	3995562824010	2064373319844	6059936143854
2	2015	2013597591230	2081854303667	4095451894897	2115982652996	6211434547893

Dossier 2 : Modélisation d'ESN

Présentation du contexte

Consciente de l'importance de modifier nos habitudes nutritionnelles et de protéger notre environnement, la société ESN a développé l'enseigne Espace Santé Nature qui offre une large gamme de produits issus de l'agriculture biologique, labellisés et contrôlés par des organismes agréés.

Le réseau commercial des boutiques Espace Santé Nature offre des produits alimentaires de consommation courante certifié , mais également des compléments alimentaires.

Ces produits, dits « nutraceutiques », ont un effet physiologique bénéfique ou protecteur contre les maladies chroniques. Ils sont fabriqués à partir de substances alimentaires et sont commercialisés sous forme de comprimé, de poudre, de potion ou d'autres formes médicinales.

Par ailleurs, l'enseigne propose des produits d'entretien naturels pour la maison (sans produits chimiques), des cosmétiques biologiques (crème de jour et nuit, gel douche, shampoing) et de nombreux éco-produits tels que de la vaisselle compostable, des couches pour bébés biodégradables à 100 % et non blanchies.

Encouragée par l'essor du marché des nutraceutiques, la société a décidé de faire évoluer son système d'information pour accompagner le développement de son enseigne Espace Santé Nature.

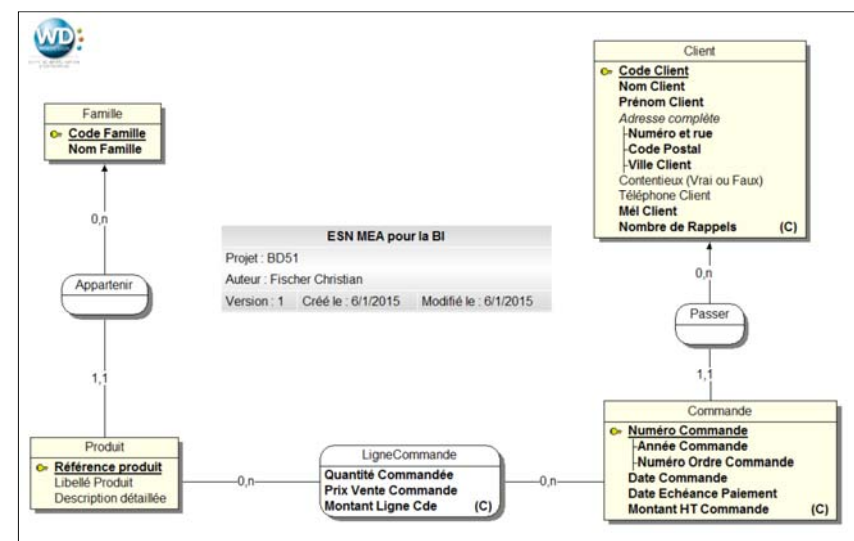
La société désire aujourd'hui adapter son site www.espace-sante-nature.com afin de commercialiser ses nouveaux produits nutraceutiques.

Le développement d'un site dynamique est basé sur une architecture trois tiers. Cette architecture nécessite l'accès à la base de données qui gère les stocks et la commercialisation des produits.

La mise en place d'un outil de reporting devra permettre de suivre les ventes par jour (à J+1 les données de ventes du site sont remontées dans un entrepôt de données).

Le suivi des ventes sera effectué par jour, semaine, mois, trimestre et année. Les périodes promotionnelles doivent également être gérées.

Modèle entité association de la base de données actuelle



Travail à faire :

1. A partir du MEA de la base actuelle, fournir un modèle en étoile utilisé par l'application décisionnelle.
2. Définir la structure de la table du calendrier des ventes et expliquer le principe de sa construction automatique.
3. Proposer une solution logicielle pour mettre en œuvre la solution de reporting décisionnelle.