

BD51

Business Intelligence & Data Warehouse

Examen Final du 15/1/2013

Rédacteur : Christian FISCHER

Département Informatique

Automne 2012

Conditions de réalisation :

Documents autorisés : Supports de Cours et TP de BD51

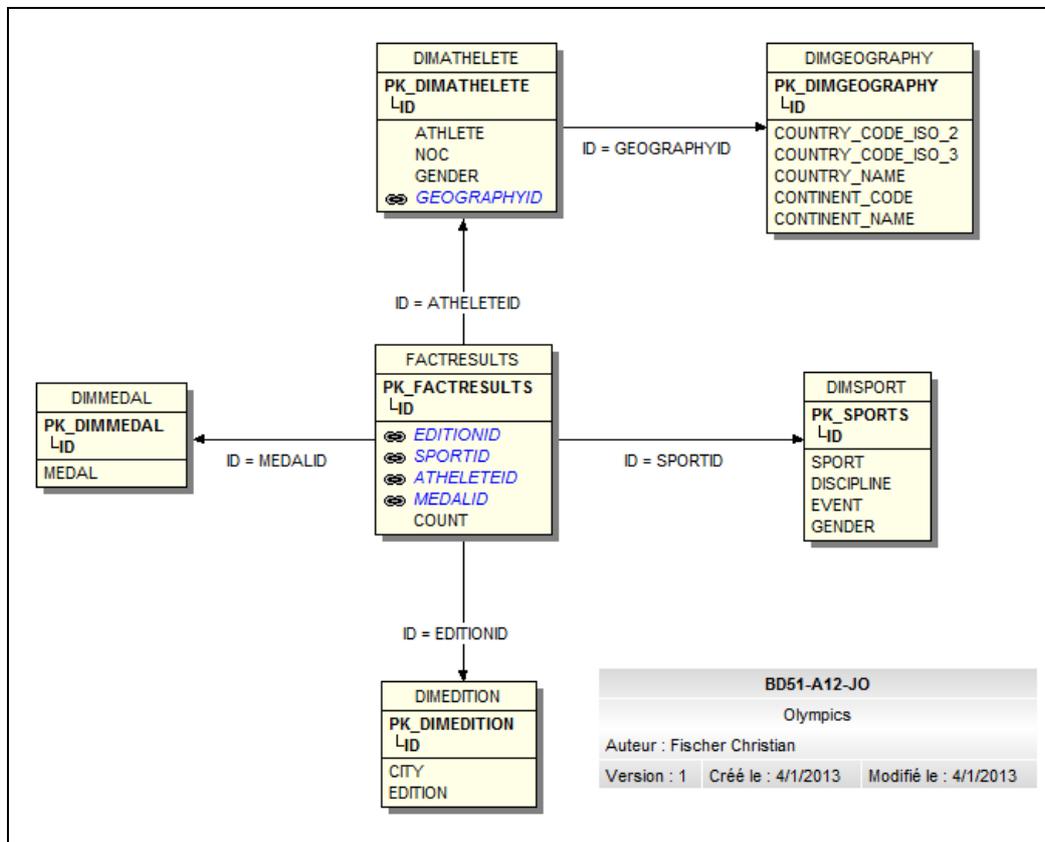
Durée : 2h

Barème : Le sujet comprend 2 dossiers indépendants.

Dossier 1 10 pts	1. Requêtes SQL et fonctions OLAP	10 pts
Dossier 2 10 pts	Modélisation du Data Warehouse 1. Modèle dimensionnel des entraînements de natation 2. Modèle dimensionnel des compétitions de triathlon Q2 : Q3 :	4 pts 4 pts 2 pts

Dossier 1 : Interrogation du système d'information des jeux olympiques

Le modèle dimensionnel en flocon de la base de données OLYMPICS pour le suivi des résultats aux jeux olympiques de 1896 à 2008 est le suivant :



Contenu des tables de dimensions

Table des médailles (DIMMEDAL)

ID	Medal
1	Gold
2	Silver
3	Bronze

Tables des éditions des jeux olympiques (DIMEDITION)

ID	City	Edition
1	Athens	1896
2	Paris	1900
3	St Louis	1904
4	London	1908
5	Stockholm	1912
6	Antwerp	1920
7	Paris	1924
8	Amsterdam	1928
9	Los Angeles	1932
10	Berlin	1936
11	London	1948
12	Helsinki	1952
13	Melbourne / Stockholm	1956
14	Rome	1960
15	Tokyo	1964
16	Mexico	1968
17	Munich	1972
18	Montreal	1976
19	Moscow	1980
20	Los Angeles	1984
21	Seoul	1988
22	Barcelona	1992
23	Atlanta	1996
24	Sydney	2000
25	Athens	2004
26	Beijing	2008

Extrait de la table des sports (DIMSPORTS)

ID	Sport	Discipline	Event	Gender
1	Aquatics	Diving	10m platform	M
2	Aquatics	Diving	10m platform	W
3	Aquatics	Diving	3m springboard	M
4	Aquatics	Diving	3m springboard	W
5	Aquatics	Diving	plain high diving	M
6	Aquatics	Diving	plunge for distance	M
7	Aquatics	Diving	synchronized diving 10m platform	M
8	Aquatics	Diving	synchronized diving 10m platform	W
9	Aquatics	Diving	synchronized diving 3m springboard	M
10	Aquatics	Diving	synchronized diving 3m springboard	W
11	Aquatics	Swimming	100m backstroke	M
12	Aquatics	Swimming	100m backstroke	W
13	Aquatics	Swimming	100m breaststroke	M
14	Aquatics	Swimming	100m breaststroke	W
15	Aquatics	Swimming	100m butterfly	M
16	Aquatics	Swimming	100m butterfly	W
17	Aquatics	Swimming	100m freestyle	M
18	Aquatics	Swimming	100m freestyle	W
19	Aquatics	Swimming	100m freestyle for sailors	M
20	Aquatics	Swimming	1200m freestyle	M

Extrait de la table des athlètes (DIMATHLETE)

ID	Athlete	NOC	Gender	GeographyID
1	AABYE, Edgar	ZZX	M	NULL
2	AALTONEN, Arvo Ossian	FIN	M	71
3	AALTONEN, Paavo Johannes	FIN	M	71
4	AAMODT, Ragnhild	NOR	W	168
5	AANING, Alf Lied	NOR	M	168
6	AARDENBURG, Willemien	NED	W	NULL
7	AARDEWIJN, Pepijn	NED	M	NULL
8	AARONES, Ann Kristin	NOR	W	168
9	AAS, Karl Jan	NOR	M	168
10	AAS, Thomas Valentin	NOR	X	168
11	ABAJO, Jose Luis	ESP	M	69
12	ABAKUMOVA, Maria	RUS	W	192
13	ABALMASAU, Aliaksei	BLR	M	NULL
14	ABALO, Luc	FRA	M	76
15	ABANDA ETONG, Patrice	CMR	M	48
16	ABARCA, Jose Maria	ESP	M	69

Tous les liens avec les pays ne sont pas renseignés dans l'exemple ci-dessus dans la colonne GeographyID : ne pas tenir compte de cette situation dans l'écriture des requêtes.

Extrait de la table des pays et continents (DIMGEOGRAPHY)

ID	Country Code ISO 2	Country Code ISO 3	Country Name	Continent Code	Continent Name
72	FJ	FJI	Fiji	OC	Oceania
73	FK	FKK	Falkland	SA	South America
74	FM	FMM	Micronesia	OC	Oceania
75	FO	FOO	Faroe	EU	Europe
76	FR	FRA	France	EU	Europe
77	GA	GAB	Gabon	AF	Africa
78	GB	GBR	United King...	EU	Europe
79	GD	GDD	Grenada	NA	North America
80	GE	GEO	Georgia	EU	Europe
81	GF	GFF	French	SA	South America
82	GG	GGY	Guemsey	EU	Europe
83	GH	GHA	Ghana	AF	Africa
84	GI	GIB	Gibraltar	EU	Europe
85	GL	GLL	Greenland	NA	North America
86	GM	GMB	Gambia	AF	Africa
87	GN	GNN	Guinea	AF	Africa
88	GP	GPP	Guadeloupe	NA	North America
89	GQ	GQQ	Equatorial G...	AF	Africa
90	GR	GRC	Greece	EU	Europe
91	GS	GSS	South	AN	Antarctica
92	GT	GTM	Guatemala	NA	North America

Extrait de la table des faits FACTRESULTS

ID	EditionID	SportID	AtheleteID	MedalID	Count
1	1	17	7069	1	1
2	1	17	7655	2	1
3	1	19	4664	3	1
4	1	19	11873	1	1
5	1	19	3011	2	1
6	1	20	3176	3	1
7	1	20	7069	1	1
8	1	20	461	2	1
9	1	36	3176	3	1
10	1	36	13608	1	1
11	1	36	14741	2	1
12	1	96	10580	3	1
13	1	96	18699	3	1
14	1	96	2516	1	1
15	1	96	7820	2	1
16	1	100	3738	1	1
17	1	100	6590	2	1
18	1	101	10978	3	1
19	1	101	5524	1	1
20	1	101	1711	2	1
21	1	117	6389	3	1
22	1	117	2516	1	1

La colonne COUNT contient toujours la valeur 1.

La granularité des tables des faits correspond à l'attribution d'une médaille à un athlète pour un sport lors d'une olympiade.

Travail à faire : Rédiger les requêtes SQL ci-dessous.

Utilisation des jointures ANSI, des alias de tables (première lettre de la table) et des alias de colonnes (nom entre parenthèse)

Requête 1 : (0,5 pt)

Calculer le nombre de médailles dans la table des faits

Résultat

Nb médailles
29216

Requête 2 : (1 pt)

Calculer le nombre de médailles par édition des JO. Tri décroissant sur l'édition.

Extrait du résultat

Edition	nb médaillés
2008	2042
2004	1998
2000	2015
1996	1859
1992	1705
1988	1546
1984	1459
1980	1387
1976	1305
1972	1185

Requête 3 : (1 pt)

Calcul du nombre de médailles pour tous les pays européens. Tri décroissant sur le nombre de médailles puis croissant sur le nom de pays en cas d'égalité.

Le nombre de médaille sera égal à 0 pour les pays sans médaille.

Extrait des 55 pays à afficher.

Pays	Nb médaillés
United Kingdom	1594
France	1314
Italy	1228
Hungary	1053
Romania	624
Norway	537
Finland	451
Belgium	408
Spain	377
Lithuania	50
Czech	41
Bosnia	23
Georgia	18
Iceland	17
Azerbaijan	16
Moldova	6
Slovenia	4
Luxembourg	2
Macedonia	1
Åland	0
Albania	0
Andorra	0
Ameria	0

Requête 4 : (1 pt) Calculer le nombre de médailles par couleur.

Le résultat sera trié par ordre de couleur : Or (Gold), Argent (Silver) et Bronze en utilisant un Case dans la clause Order by.

Résultat

Couleur médaille	Nb médaillés
Gold	9850
Silver	9677
Bronze	9689

Requête 5 : (1 pt)

Pour les éditions à partir de 2000 calculer le nombre de médaillés

Le résultat sera trié par édition décroissante et par ordre de couleur : Or (Gold), Argent (Silver) et Bronze (au choix tri du ID de la médaille ou sur le nom de la couleur comme dans la requête 4)

Résultat

Année	Ville	Couleur	Nb médaillés
2008	Beijing	Gold	669
2008	Beijing	Silver	663
2008	Beijing	Bronze	710
2004	Athens	Gold	659
2004	Athens	Silver	660
2004	Athens	Bronze	679
2000	Sydney	Gold	663
2000	Sydney	Silver	667
2000	Sydney	Bronze	685

Requête 6 : (1 pt)

Calcul du nombre de médailles par pays et couleur de médaille

Ajout du nombre de médailles total par pays (grouping sets).

Tri du résultat par nom pays décroissant et par couleur de médaille

Extrait du résultat

Country Name	Couleur	Nombre de médailles
USA	Gold	2088
USA	Silver	1195
USA	Bronze	1052
USA	Total Pays	4335
United Kingdom	Gold	498
United Kingdom	Silver	591
United Kingdom	Bronze	505
United Kingdom	Total Pays	1594
Spain	Silver	193
Spain	Total Pays	377
Russian	Gold	192
Russian	Silver	206
Russian	Bronze	240
Russian	Total Pays	638
Romania	Gold	155

Requête 7 : (1 pt)

Calcul du nombre de médailles par édition obtenus par les athlètes Français.

Ajout du total par édition dans la dernière colonne.

Tri du résultat par édition décroissante et par couleur de médaille à l'aide de colonne ID.

Extrait du résultat

Edition	City	ID	Medal	Nombre de médailles françaises	Nb médaillés par édition
2008	Beijing	1	Gold	25	76
2008	Beijing	2	Silver	23	76
2008	Beijing	3	Bronze	28	76
2004	Athens	1	Gold	21	53
2004	Athens	2	Silver	10	53
2004	Athens	3	Bronze	22	53
2000	Sydney	1	Gold	22	66
2000	Sydney	2	Silver	30	66
2000	Sydney	3	Bronze	14	66
1996	Atlanta	1	Gold	21	51
1996	Atlanta	2	Silver	10	51
1996	Atlanta	3	Bronze	20	51
1992	Barcelona	1	Gold	9	57
1992	Barcelona	2	Silver	5	57

Requête 8 : (1,5 pt) Liste de médaillés Français de l'édition 2008

Utilisation de ROW_NUMBER pour les colonnes 1 et 6

Dernière colonne calcul du nombre de médailles par sport.

Tri du résultat par sport, athlète et couleur de médaille.

Extrait du résultat (76 médaillés)

Numéro Ligne	Sport	Athlete	ID	Medal	Numéro médaillé par sport	Nb médaillés par sport
1	Aquatics	BERNARD, Alain	1	Gold	1	11
2	Aquatics	BERNARD, Alain	2	Silver	2	11
3	Aquatics	BERNARD, Alain	3	Bronze	3	11
4	Aquatics	BOUSQUET, Frederick	2	Silver	4	11
5	Aquatics	DUBOSCQ, Hugues	3	Bronze	5	11
6	Aquatics	DUBOSCQ, Hugues	3	Bronze	6	11
7	Aquatics	GILLOT, Fabien	2	Silver	7	11
8	Aquatics	LEVEAUX, Amaury	2	Silver	8	11
9	Aquatics	LEVEAUX, Amaury	2	Silver	9	11
10	Aquatics	MALLET, Gregory	2	Silver	10	11
11	Aquatics	STEIMETZ, Boris	2	Silver	11	11
12	Archery	ARNOLD, Virginie	3	Bronze	1	3
13	Archery	DODEMONT, Sophie	3	Bronze	2	3
14	Archery	SCHUH, Berengere	3	Bronze	3	3
15	Athletics	BAALA, Mehdi	3	Bronze	1	2
16	Athletics	MEKHISSI-B., Mahie...	2	Silver	2	2
17	Boxing	DJELKHIR, Khedafi	2	Silver	1	3
18	Boxing	SOW, Daouda	2	Silver	2	3
19	Boxing	VASTINE, Alexis	3	Bronze	3	3

Requête 9 : (2 pts) Utilisation de l'opérateur PIVOT

Afficher l'édition, la ville le nombre de médailles en Or, Argent, et Bronze et le nombre total de médailles gagnés par la France.

Tri de résultat par Edition décroissante.

Extrait du résultat

Edition	Ville	Or	Argent	Bronze	Total
2008	Beijing	25	23	28	76
2004	Athens	21	10	22	53
2000	Sydney	22	30	14	66
1996	Atlanta	21	10	20	51
1992	Barcelona	9	5	43	57
1988	Seoul	12	4	13	29
1984	Los Angeles	21	16	31	68
1980	Moscow	18	5	6	29
1976	Montreal	5	7	8	20
1972	Munich	2	5	15	22
1968	Mexico	12	5	10	27
1964	Tokyo	1	13	17	31
1960	Rome	0	6	5	11
1956	Melbourne / Stockholm	6	13	13	32
1952	Helsinki	14	9	16	39
1948	London	25	22	21	68
1936	Berlin	12	13	18	43
1932	Los Angeles	23	8	6	37
1928	Amsterdam	7	16	13	36
1924	Paris	39	63	20	122
1920	Antwerp	13	73	55	141

Dossier 2 A : Modèle dimensionnel des entraînements de natation

Le Comité Départemental de Natation du Territoire de Belfort (CDNTB) organise des entraînements de natation communs aux différents clubs de l'aire urbaine dans le but d'harmoniser les pratiques et de déceler les futurs talents.

Ces entraînements communs nécessitent de disposer de créneaux horaires dans trois piscines différentes du département.

Une piscine se caractérise par un numéro, un nom, une adresse complète ainsi que la longueur du bassin (25 ou 50 mètres).

Un entraînement est repéré par un numéro d'entraînement, une date, une heure de début et une heure de fin et se déroule dans une piscine.

Le président du CDNTB souhaite mettre en place une gestion informatisée afin de contrôler que chaque athlète suit bien son plan d'entraînement personnalisé.

Un athlète est identifié par un numéro de licence, un nom et prénom, une date de naissance, une catégorie (minime, benjamin, junior, senior ou vétéran), un sexe (H ou F).

Pour chaque athlète, le plan d'entraînement personnalisé permet de planifier la distance à parcourir pour chaque entraînement.

A l'issue de chaque entraînement la distance parcourue est notée et un test chronométré est effectué sur la distance de 100 m nage libre pour suivre l'évolution des performances hors compétition de chaque athlète.

Travail à faire

Q1. Fournir un modèle dimensionnel en étoile.

Dans le modèle fourni vous indiquerez quelles sont les tables de dimensions et la table de faits.

Dossier 2 B : Modèle dimensionnel des compétition de triathlon

Le triathlon, né dans l'archipel d'Hawaï il y a vingt ans, discipline olympique en ouverture des jeux de Sydney en septembre 2000, est un sport jeune en cours de développement.

La discipline, lancée par le capitaine de vaisseau John Collins, consiste dans l'enchaînement de trois sports de base dans un ordre déterminé immuable : natation (issue du « Waikiki rough water swim » entre deux îles d'Hawaï), cyclisme sur route (issu du « Around Oahu bike race », course cycliste de l'île Oahu), course à pied (inspirée du marathon d'Honolulu).

Le Comité Départemental de Natation du Territoire de Belfort (CDNTB) se charge également de l'organisation des épreuves qui se déroulent dans son département et sont ouvertes aux triathlètes de la France entière.

Les triathlètes (« équivalent aux athlètes du dossier 2A »)

Chaque triathlète possède une licence délivrée par la Fédération Française de Triathlon. Le numéro de licence identifie un sportif et reste identique d'une année à l'autre.

Un triathlète possède également un nom et prénom, une date de naissance, une catégorie (minime, benjamin, junior, senior ou vétéran), un sexe (H ou F).

Les triathlons

Un triathlon est une compétition ouverte à toutes les catégories d'âge.

Il existe plusieurs types de triathlon, chaque type de triathlon se distingue des autres par les distances à parcourir dans les trois sports de base (NAT=natation, CYC=cyclisme et CAP=course à pied).

Chaque type de triathlon est identifié par un code et possède une désignation.

- LD pour longue distance. (ou IronMan) : 3800 m NAT, 180 km CYC et 42,195 km CAP
- DO pour distance olympique : 1500m NAT, 40 km CYC, 10km CAP
- ...

Chaque triathlon possède un numéro, un nom, un type et se déroule en un lieu précis à une date donnée.

Par exemple, le triathlon numéro 12, nommé « Course folle », se déroule au lac du Malsaucy le 26 juin 2012; il s'agit d'une compétition de type DO.

Inscription et résultats

Dès que l'organisation d'un triathlon a été publiée par le CDNTB, les triathlètes peuvent s'y inscrire.

À l'inscription, un numéro de dossard est attribué au triathlète pour la compétition et sa date d'inscription est mémorisée. Une inscription est identifiée par le numéro du triathlon concerné suivi du numéro de dossard attribué au triathlète.

Les inscriptions sont closes 15 jours avant la date de la compétition.

Le jour du triathlon, le CDNTB recense les participants à la course. Il arrive qu'un sportif inscrit ne se présente pas, il est alors déclaré forfait.

Après le déroulement de la course, les résultats obtenus par chaque participant sont relevés :

- Temps réalisé dans les différentes épreuves (natation, course cycliste, course à pied).
- Temps total réalisé pour le triathlon
- Classement du concurrent dans sa catégorie

Le cas des abandons en cours d'épreuve sort du cadre de cette étude et n'est donc pas à prendre en compte.

Travail à faire

Q2. Fournir un modèle dimensionnel en étoile permettant de mesurer les performances des athlètes qui participent aux triathlons.

Dans le modèle fourni vous indiquerez quelles sont les tables de dimensions et la table de faits en indiquant son niveau de granularité.

Q3. Indiquez les autres dimensions ou axes d'analyse susceptibles d'être ajoutés au modèle.