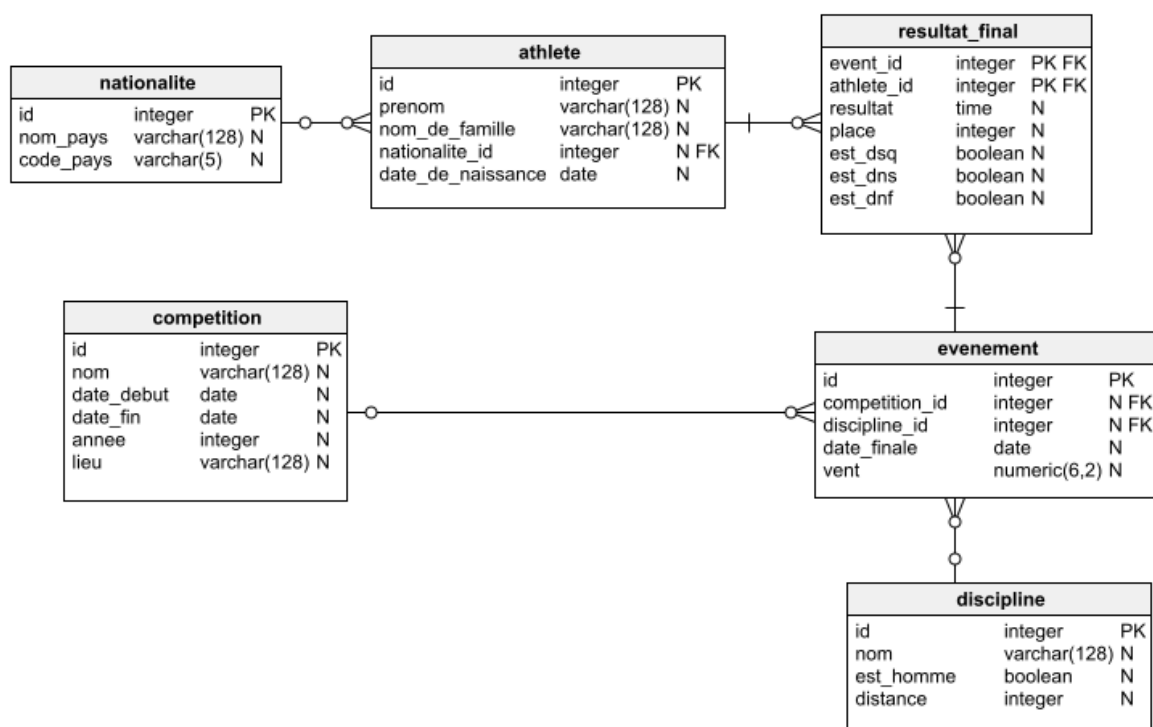


Requêtes SQL : Sélection, agrégation et regroupement

Objectifs	Révision sur les sélections simples à savoir : <ul style="list-style-type: none"> • La sélection de lignes • La sélection de colonnes • Ordonner les lignes • Renommer les colonnes • Limiter le nombre de lignes • Supprimer les doublons • Grouper
Méthodologie	Travail individuel
Durée estimée	45 min (avec le corrigé)
Type de base de données :	MySQL

Dans cet exercice, nous allons travailler avec des données concernant les finales de différentes disciplines de course sur piste lors de quelques championnats d'athlétisme : Jeux Olympiques de Rio de Janeiro en 2016, Championnats du monde d'athlétisme de l'IAAF de Londres en 2017, et Championnats du monde d'athlétisme de l'IAAF de Doha en 2019.

Jetons un coup d'œil aux données que nous allons traiter.



La base de données comporte six tables :

competition, **evenement** , **discipline**, **resultat_final**, **athlete** et **nationalite**.

Passons en revue chacune d'entre elles.

La table **competition** comporte les colonnes suivantes :

id - un identifiant.

nom - le nom de la compétition.

date_debut - le premier jour de la compétition.

date_fin - le dernier jour de la compétition.

annee - l'année au cours de laquelle une compétition donnée a eu lieu.

lieu - le lieu où s'est déroulée une compétition donnée.

Voici quelques exemples de lignes :

id	nom	date_debut	date_fin	annee	lieu
7093747	Jeux Olympiques de Rio de Janeiro	2016-08-12	2016-08-21	2016	Stade olympique, Rio de Janeiro (BRA)
7093740	Championnats du monde d'athlétisme de l'IAAF de Londres	2017-08-04	2017-08-13	2017	Stade olympique, Londres (GBR)
7125365	Championnats du monde d'athlétisme de l'IAAF de Doha	2019-09-27	2019-10-06	2019	Stade international Khalifa, Doha (QAT)

La table **discipline** comporte les colonnes suivantes :

id - un identifiant.

nom - le nom de la discipline donnée.

est_homme - TRUE s'il s'agit d'une discipline masculine, FALSE s'il s'agit d'une discipline féminine.

distance - la distance de la discipline donnée, exprimée en mètres.

Nous stockons des informations pour toutes les disciplines de la course à pied - des courses de courte distance comme le 100 mètres aux courses de longue distance comme le marathon.

Voici quelques exemples de lignes :

id	nom	est_homme	distance
1	Hommes 100m	true	100
16	Marathon femmes	false	42195

La table **evenement** stocke les informations relatives à un événement, par exemple "100m femmes", et comporte les colonnes suivantes :

id - un identifiant.

competition_id - correspondant à l'identifiant de la compétition.

discipline_id - correspondant à l'identifiant de la discipline.

date_finale - la date à laquelle s'est déroulée la finale d'une épreuve donnée.

vent - les points de vent pendant la finale.

La table **athlete** comporte les colonnes suivantes :

id - un identifiant.

prenom - le prénom de l'athlète.

nom_de_famille - le nom de famille de l'athlète

nationalite_id - la nationalité de l'athlète.

date_de_naissance - la date de naissance de l'athlète.

La table **nationalite** stocke les informations relatives aux pays et comporte les colonnes suivantes :

id - un identifiant.

nom_pays - le nom du pays.

code_pays - l'abréviation à trois lettres du pays.

La table **resultat_final** contient des informations sur tous les participants à un événement particulier :

event_id - l'identifiant d'un événement correspondant.

athlete_id - l'identifiant d'un athlète.

resultat - le temps réalisé par un athlète donné. Peut être manquant.

place - la place d'un athlète donné. Peut être manquant.

est_dsq - TRUE en cas de disqualification.

est_dnf - TRUE si un athlète donné n'a pas terminé la course (did not finish).

est_dns - TRUE si un athlète donné n'a pas pris le départ de la course (did not start).

Les trois dernières colonnes peuvent prêter à confusion, mais cette notation est largement utilisée. Par exemple, voici à quoi ressembleraient les résultats du 5000 mètres masculin en 2019, par exemple, à la télévision :

1	Muktar EDRIS	ETH	00:12:58.85
2	Selemon BAREGA	ETH	00:12:59.7
3	Mohammed AHMED	CAN	00:13:01.11
4	Telahun Haile BEKELE	ETH	00:13:02.29
5	Jakob INGEBRIGTSEN	NOR	00:13:02.93
6	Jacob KROP	KEN	00:13:03.08
7	Paul CHELIMO	USA	00:13:04.6
8	Nicholas Kipkorir KIMELI	KEN	00:13:05.27
9	Birhanu BALEW	BRN	00:13:14.66
10	Justyn KNIGHT	CAN	00:13:26.63
11	Hassan MEAD	USA	00:13:27.05
12	Stewart MCSWEYN	AUS	00:13:30.41
13	Henrik INGEBRIGTSEN	NOR	00:13:36.25
14	Isaac KIMELI	BEL	00:13:44.29
-	Filip INGEBRIGTSEN	NOR	DNF

Muktar Edris s'est imposé avec le résultat de 12 minutes et 58,85 secondes, son meilleur résultat de la saison.

L'un des frères Ingebrigtsen - Filip - n'a pas terminé la course. A un moment donné, il est sorti de la piste. Le statut de son résultat est DNF - n'a pas terminé.

Les autres statuts possibles sont :

DNS - n'a pas pris le départ

DSQ - disqualifié.

Les indicateurs booléens `est_dnf`, `est_dns`, et `est_dsq` correspondent à ces statuts possibles.

Voici ce que cela donne dans notre base de données, dans la table **resultat_final**:

event_id	athlete_id	resultat	place	est_dsq	est_dns	est_dnf
40	14464221	00:12:58.85	1	false	false	false
40	14751317	00:12:59.7	2	false	false	false
40	14330105	00:13:01.11	3	false	false	false
40	14797485	00:13:02.29	4	false	false	false
40	14653717	00:13:02.93	5	false	false	false
40	14808798	00:13:03.08	6	false	false	false
40	14454958	00:13:04.6	7	false	false	false
40	14708978	00:13:05.27	8	false	false	false
40	14695983	00:13:14.66	9	false	false	false
40	14603387	00:13:26.63	10	false	false	false
40	14245833	00:13:27.05	11	false	false	false
40	14500488	00:13:30.41	12	false	false	false
40	14215338	00:13:36.25	13	false	false	false
40	14482919	00:13:44.29	14	false	false	false
40	14458705	NULL	NULL	false	false	true

Pour obtenir les noms, dates de naissance et nationalités des athlètes, vous devrez utiliser d'autres tables. Comme vous pouvez le constater, la dernière ligne correspond à une course non terminée, de sorte que les valeurs de temps et de place sont manquantes.

Requête n°1 :

Où se sont déroulées toutes les compétitions ? Indiquez tous les lieux.

Requête n°2 :

Affichez les dates finales de tous les événements et les points de vent.

Requête n°3 :

Sélectionnez tous les noms et prénoms, ainsi que les dates de naissance de tous les athlètes.

Requête n°4 :

Sélectionnez tous les noms de compétitions ainsi que leurs dates de début et de fin.

Requête n°5 :

Sélectionnez toutes les colonnes pour les lignes dont le résultat est en première position.

Requête n°6 :

Affichez toutes les données relatives aux compétitions qui se sont déroulés avant l'année 2018.

Requête n°7 :

Affichez tous les résultats (temps) pour lesquels la place était égale ou supérieure à 100.

Requête n°8 :

Affichez toutes les finales où le vent était supérieur à 0,5.

Requête n°9 :

Affichez tous les noms de disciplines dont la distance est supérieure à 3000 mètres.

Requête n°10 :

Affichez les noms et prénoms des athlètes dont le prénom commence par K et se termine par a.

Requête n°11 :

Affichez tous les noms de pays commençant par 'U'.

Requête n°12 :

Affichez tous les noms de pays composés d'au moins trois mots.

Requête n°13 :

Quel est le pays dont l'abréviation est 'ART'? Affichez son nom.

Requête n°14 :

Affichez toutes les lignes de la table discipline relatives aux marathons (contenant 'Marathon' dans le nom de la discipline).

Requête n°15 :

Affichez toutes les lignes de resultat_final où le coureur a terminé sans place.

Requête n°16 :

Affichez toutes les lignes de la table resultat_final où une disqualification s'est produite.

Requête n°17 :

Affichez toutes les lignes de la table resultat_final où le coureur n'a pas pris le départ.

Requête n°18 :

Affichez tous les noms de disciplines féminines.

Requête n°19 :

Affichez toutes les lignes de la table resultat_final où un coureur a été disqualifié mais a commencé et terminé la course.

Requête n°20 :

Affichez toutes les lignes de resultat_final pour tous les médaillés d'or dont le temps est supérieur à une heure..

Requête n°21 :

Affichez tous les noms des disciplines masculines si la distance est inférieure à 500 mètres.

Requête n°22 :

Affichez tous les noms et prénoms des athlètes s'ils contiennent un 'z' (un ou plusieurs, en minuscules ou en majuscules) dans leur nom et leur prénom.

Requête n°23 :

Indiquez le prénom et le nom de tous les athlètes s'ils contiennent un 'x' (un ou plusieurs, en minuscules ou en majuscules) dans l'un ou l'autre de leurs noms.

Requête n°24 :

Affichez les noms et abréviations des pays. Trier les lignes par nom de pays, dans l'ordre croissant.

Requête n°25 :

Affichez tous les noms des disciplines féminines possibles, triés par la distance, dans l'ordre décroissant.

Requête n°26 :

Affichez tous les temps et places des résultats. Trier les temps par ordre croissant. Si plusieurs résultats sont identiques, affichez d'abord celui qui a la meilleure place.

Requête n°27 :

Pour tous les athlètes, indiquez leur nom et leur prénom. Triez par le nom de famille, et si quelques athlètes le partagent, triez par leur prénom (les deux dans l'ordre croissant).

Requête n°28 :

Affichez le nom et le prénom de tous les athlètes dont le nom de famille commence par 'X'. Triez le résultat par le prénom dans l'ordre décroissant.

Requête n°29 :

Pour tous les résultats finaux, affichez les résultats de temps qui sont d'au moins 3 heures. Triez les lignes par ordre décroissant des résultats.

Requête n°30 :

Affichez tous les noms de pays commençant par 'T' ou 'Z' triés par ordre croissant..

Requête n°31 :

Affichez toutes les disciplines masculines triées par distance dans l'ordre décroissant. Affichez deux colonnes : nom et distance.

Requête n°32 :

Sélectionnez toutes les lignes pour tous les événements où les points de vent étaient positifs. Trier par ordre décroissant des points de vent.

Requête n°33 :

Comptons le nombre de nationalités que nous avons dans notre base de données. Affichez une colonne : compte_nationalite.

Requête n°34 :

Afficher le nombre d'événements que nous stockons et renommer la colonne : compte_finales.

Requête n°35 :

Voyons combien de courses se sont terminées avec et sans résultat. Affichez deux colonnes : **courses_reussies** et **courses_non_reussies**.

Requête n°36 :

Cherchons le prénom le plus populaire. Affichez deux colonnes : **prenom** et **nombre_de_noms**. Triez les résultats par ordre décroissant des prénoms..

Requête n°37 :

Trouvons le prénom le plus populaire.

Affichez deux colonnes : **prenom** et **nombre_de_noms**, mais uniquement si le nombre est au moins égal à 5.

Triez le résultat par ordre décroissant.

Requête n°38 :

Comptons combien il y a de pays qui contient 'Républiques' dans leur nom. N'affichez qu'une seule colonne : **compte_republiques**.

Requête n°39 :

Affichez tous les noms de disciplines féminines.

Vous devez utiliser la fonction EXTRACT() pour extraire l'année :

EXTRACT(YEAR FROM date_de_naissance)