Question 1 : Un diagramme entité-relation, également nommé diagramme ER est un schéma graphique avec les entités et les relations liant ces entités. Pour réaliser ce schéma. Pour réaliser ce diagramme, mon binôme et moi avons listé les entités listées dans l’énoncé puis les regrouper selon une entité globale qui détient des caractéristiques (ex : Attaque, Réponse, Agent…). Puis on a déterminé comment les entités interagissent entre elles et indiqué la cardinalité entre les relations. On a enfin relu le schéma pour voir s’il n’y avait pas de redondance et s’il était optimisé à la recherche d’informations.

Pour passer de ce schéma à un modèle physique 3NF, on a pour chaque entité, créez une table dans le modèle physique puis nommé les tables de manière descriptive et utiliser le nom de l’entité correspondante. Pour chaque table, on a défini les colonnes basées sur les attributs de l’entité puis spécifié les types de données pour chaque colonne (ex. : INTEGER, TEXT, etc.). Enfin on a choisi un attribut (ou un ensemble d'attributs) comme clé primaire pour chaque table, garantissant l’unicité et pour les clés étrangères on a Identifié les relations et ajouté des clés étrangères dans les tables appropriées pour établir les liens entre elles.

Question 2 :

On souhaite diviser la colonne réponse du document csv en 2 colonnes type de réponse et source de la réponse. Pour ce faire on utilise la bibliothèque panda pour lire le fichier csv et pour la colonne réponse on a repéré que les 2 informations source et type étaient séparées par un espace. Donc on divise la colonne réponse en 2 à l’endroit où il y a un espace pour créer 2 colonnes.

Question 3 :

On souhaite ajouter une colonne pour savoir si l’attaque a un groupe d’attaquants qui est confirmée ou non. On lit le fichier csv avec la bibliothèque panda. Puis on crée une colonne « Attackers confirmed » qu’on remplit par True si dans la colonne affiliations le groupe est confirmée et par False si l’enquête n’est pas sûre des attaquants.

Question 4 :

On cherche à passer de la configuration du fichier à un dictionnaire. On crée donc un dictionnaire vide appelé config. Puis on lit toutes les lignes du fichier avec une boucler et la fonction readline. A la fin de toutes les lignes on lui affecte la clé et la valeur correspondant à la configuration du fichier. Cette affectation est dans la boucle car l’algorithme doit lire toutes les lignes pour avoir la configuration du fichier.

Question 5 :

Pour cette question il faut créer notre base de données avec nos colonnes trouvées dans la question 1. Pour ce faire, on crée un dictionnaire appelé tables qui permet de mettre toutes nos colonnes en chaine de caractère avec leur caractéristiques définis avec leur type (INTEGER,TEXT,BINARY) ou encore si elles sont des clés primaires, étrangères.

Question 6 : Pour mettre le pays des attaquants dans la table attaque on suppose que les attaquants proviennent du pays du sponsor. Donc une seule colonne sponsor indiquant le pays sponsorisant et la provenance des attaquants.

Pour cette question on vérifie si la base de données est bien peuplée donc on vérifie les données si elles respectent bien les caractéristiques énoncée dans la question 5 et renvoyant donc un booléen.