



NIVEAU 2 SUR 3

Standard

Le niveau d'un contrôle de classe : vérifier un fichier, enchaîner des requêtes, analyser un cas.

Questions guidées, sans banque de mots · 60 minutes · sur 20

Données — Structuration et traitement

Corrigé du contrôle **niveau 2 · Standard**, avec le barème au point près.

📖 RAPPEL

La règle du cours que l'exercice met à l'épreuve, en tête de chaque exercice.

✓ RÉPONSE

Dans le cadre bordé : ce qui rapporte les points.

→ POURQUOI

Sous chaque réponse : la raison, pour comprendre et non recopier.

1

Vérifier avant de calculer

⌚ Vérifier un fichier · Traiter une table

⌚ 11 min

4 pts

RAPPEL DE COURS

📖 Exploiter un fichier commence par l'identification du format et du séparateur, puis la lecture de l'en-tête. On vérifie ensuite le type de chaque champ et l'on repère les valeurs manquantes ou aberrantes : un nombre stocké comme texte fausse tri et calcul, une case vide n'est pas un zéro. On travaille sur une copie, en gardant le fichier d'origine intact.

a Quel caractère joue ici le rôle de séparateur ? Combien d'élèves ce fichier décrit-il ?

1 pt

✓ **Le point-virgule ; sept élèves, un par ligne, de la ligne 2 à la ligne 8.**

→ **Pourquoi.** Le séparateur s'identifie dès la première ligne ; la ligne 1 nomme les colonnes, elle n'est pas un enregistrement. 0,5 point par réponse.

b Ligne 8, on lit « 0,5 ». Que peut signifier ce nombre, et que faudrait-il savoir pour trancher ?

1 pt

✓ **Probablement une demi-heure, quand les autres durées semblent données en minutes. Pour trancher, il faut savoir comment la durée a été demandée aux élèves : cette information doit accompagner le fichier.**

→ **Pourquoi.** Rien dans le fichier ne dit en quoi la durée s'exprime : seule une métadonnée, ou le producteur du fichier, peut trancher entre une demi-heure et une demi-minute.

c Ligne 7, la durée est écrite « vingt ». Quel problème cela pose-t-il pour un programme, et comment le traiter ?

1 pt

✓ **C'est du texte et non un nombre : un programme ne peut ni l'additionner ni le trier avec les autres durées. On la remplace par 20 dans une copie du fichier, en gardant l'original intact.**

→ **Pourquoi.** Un nombre stocké comme texte fausse tout tri et tout calcul ; travailler sur une copie permet de revenir à la donnée d'origine. 0,5 pour le problème, 0,5 pour le traitement.

d Ligne 4, la durée est vide. Calcule la durée moyenne sur les seules lignes 2, 3, 5 et 6, puis la valeur qu'on obtiendrait en comptant la case vide pour 0. Laquelle faut-il retenir ?

1 pt

✓ **$25 + 12 + 40 + 15 = 92$; $92 \div 4 = 23$ min. En comptant la case vide pour 0 : $92 \div 5 = 18,4$ min. Il faut retenir 23 min, calculée sur les valeurs connues.**

→ **Pourquoi.** Une case vide est une information non saisie, pas une durée nulle : la compter pour 0 fait baisser la moyenne sans raison. 0,5 pour les deux calculs, 0,5 pour le choix justifié.

⚠ CE QUI COÛTE DES POINTS

Une moyenne qui compte la case vide pour zéro : 0,5 sur 1.

Barème : 4 points — 1 par question.

🔍 ERREURS FRÉQUENTES

- Compter huit enregistrements en incluant la ligne d'en-tête.
- Corriger « vingt » ou « 0,5 » sans dire que le choix suppose une information que le fichier ne donne pas.
- Diviser par 5 en comptant la case vide comme une durée nulle.

2 Enchaîner des requêtes ☆ EXERCICE-PHARE

🕒 Traiter une table · Interpréter un résultat

🕒 16 min

6 pts

RAPPEL DE COURS

- Rechercher, filtrer, trier, calculer : ces quatre opérations se combinent, et l'on peut regrouper les enregistrements avant de calculer dans chaque groupe. L'ordre dans lequel on les applique change le coût du traitement, jamais le résultat. On conclut en disant ce que les données mesurent, et ce qu'elles ne mesurent pas.

a Combien de prêts concernent une classe de seconde ?

1 pt

- ✓ **Filtrer sur « classe commence par 2nde », puis calculer l'effectif : 6 prêts (P01, P03, P04, P06, P08, P09).**

→ **Pourquoi.** Un filtre suivi d'un comptage : on garde les enregistrements qui vérifient la condition, puis on les compte. 0,5 pour les opérations, 0,5 pour le résultat.

b Quels prêts ont eu lieu le mardi, du plus long au plus court ?

1 pt

- ✓ **Filtrer sur « jour = mardi », puis trier par duree_h décroissante : P03 (2 h), P05 (1,5 h), P04 (0,5 h).**

→ **Pourquoi.** Le filtre réduit la table à trois enregistrements, le tri les ordonne sans en retirer aucun.

c Pour la question précédente, Léo trie d'abord les dix prêts par durée, puis ne garde que ceux du mardi ; Sarah filtre d'abord, puis trie. Obtiennent-ils le même résultat ? Quelle démarche coûte le moins ?

1 pt

- ✓ **Le même résultat : P03, P05, P04. La démarche de Sarah coûte moins, car elle ne trie que trois lignes au lieu de dix.**

→ **Pourquoi.** L'ordre des opérations change le coût du traitement, jamais le résultat : filtrer d'abord réduit ce qu'il reste à trier.

d Regroupe les prêts par motif et calcule la durée moyenne d'un prêt pour chaque motif.

1,5 pt

- ✓ **Regrouper par motif, puis calculer la moyenne dans chaque groupe. Exposé : $1 + 2 + 1 + 0,5 + 1 = 5,5$; $5,5 \div 5 = 1,1$ h. Recherche : $1,5 + 0,5 + 1,5 + 0,5 + 2 = 6$; $6 \div 5 = 1,2$ h.**

→ **Pourquoi.** Chaque groupe compte cinq prêts : on divise la somme de chaque groupe par 5, pas par 10. 0,5 pour le regroupement, 0,5 par moyenne juste.

e Une élève conclut : « Les élèves de seconde se servent bien plus des ordinateurs que ceux de terminale. » Ce fichier permet-il de l'affirmer ? Justifie.

1,5 pt

- ✓ **Non. Le fichier compte des prêts d'une seule semaine, dans un seul CDI : il mesure la durée des prêts, pas le temps réel d'utilisation ni les ordinateurs utilisés ailleurs. Il ne dit pas non plus combien d'élèves compte chaque niveau.**

→ **Pourquoi.** C'est la dernière étape de la méthode : conclure sur ce que les données mesurent, et sur ce qu'elles ne mesurent pas. 0,5 par limite pertinente, dans la limite de 1,5.

⚠ CE QUI COÛTE DES POINTS

Un résultat juste sans les opérations qui y conduisent : la moitié des points de la question.

Barème : 6 points — 1 + 1 + 1 + 1,5 + 1,5.

🔍 ERREURS FRÉQUENTES

- Donner le résultat sans nommer les opérations qui y conduisent.
- Calculer la moyenne d'un motif en divisant par 10, le nombre total de prêts.
- Conclure sur l'usage des ordinateurs alors que le fichier ne mesure que des prêts.

3 D'une structure imbriquée à une table

📌 Structurer des données

🕒 10 min

3,5 pts

RAPPEL DE COURS



Le JSON range des couples nom-valeur et accepte des structures imbriquées ; le CSV écrit une table plate, une valeur par case. Pour passer de l'un à l'autre, on choisit ce que représente une ligne : un enregistrement par élève, par exemple.

- a Relève dans le document 3 un couple nom-valeur simple, puis un champ dont la valeur est une liste. (1 pt)

✓ **Par exemple « classe » : « 2nde B » ; le champ « accompagnateurs » (ou « élèves ») contient une liste.**

→ **Pourquoi.** Un couple simple associe un nom à une seule valeur ; les crochets signalent une liste. 0,5 point par élément.

- b Pourquoi ces données ne tiennent-elles pas telles quelles en une seule ligne de CSV ? (1 pt)

✓ **Parce que le champ « élèves » contient une liste d'objets, de longueur variable, chacun avec ses propres champs : la structure est imbriquée, alors qu'une ligne de CSV ne range qu'une valeur par case.**

→ **Pourquoi.** Le CSV convient aux tables plates, le JSON aux données arborescentes : une liste d'objets ne s'aplatit pas sans choisir une autre organisation.

- c Écris un fichier CSV qui décrit les élèves de cette sortie : une ligne d'en-tête, puis une ligne par élève. (1,5 pt)

✓ **prenom;repas;classe — Lou;pique-nique;2nde B — Sami;cantine;2nde B — Zoé;pique-nique;2nde B (quatre lignes ; on peut aussi répéter la sortie et la date sur chaque ligne).**

→ **Pourquoi.** On choisit qu'un enregistrement représente un élève ; ce qui est commun à tous, comme la classe, se répète sur chaque ligne. 0,5 pour l'en-tête, 0,5 pour les trois lignes, 0,5 pour un séparateur unique.

⚠ CE QUI COÛTE DES POINTS

Une liste écrite dans une seule case du CSV.

Barème : 3,5 points — 1 + 1 + 1,5.

🔍 ERREURS FRÉQUENTES

- Écrire toute la liste des élèves dans une seule case du CSV.
- Oublier la ligne d'en-tête, ou changer de séparateur d'une ligne à l'autre.

RAPPEL DE COURS

- Une donnée personnelle se rapporte à une personne identifiée ou identifiable, directement ou par recoupement. Le RGPD impose la finalité, la minimisation, une durée limitée de conservation et un consentement libre et éclairé ; il donne à chacun des droits d'accès, de rectification, d'effacement et d'opposition.

- a Parmi les champs des lignes 3 à 5, lesquels permettent, à eux seuls, de retrouver qui s'est inscrit ? Cites-en deux, en justifiant. (1 pt)

✓ **Par exemple la photographie et le numéro de téléphone : sans le nom, chacun suffit à reconnaître ou à joindre l'inscrit, qui reste identifiable.**

- **Pourquoi.** Une donnée est personnelle dès que la personne est identifiable, directement ou par recoupement ; le nom n'est qu'un cas parmi d'autres. 0,5 par champ justifié.

- b Au regard de l'usage annoncé (ligne 2), quels champs sont de trop ? Quel principe le formulaire méconnaît-il ? (1 pt)

✓ **La religion et la liste des sites consultés n'ont aucun rapport avec un trajet partagé (la date de naissance est discutable). Le formulaire méconnaît la minimisation.**

- **Pourquoi.** La minimisation impose de ne collecter que ce qui est nécessaire à l'usage annoncé : pour mettre en relation des élèves sur un trajet, la commune et un moyen de contact suffisent.

- c La ligne 6, puis la ligne 7 : quelle règle du RGPD chacune enfreint-elle ? (1 pt)

✓ **Ligne 6 : on ne garde les données que le temps de l'usage, jamais indéfiniment. Ligne 7 : aucun usage précis n'est annoncé, ce qui viole la finalité ; accepter un usage inconnu, ce n'est pas accepter en connaissance de cause.**

- **Pourquoi.** La conservation a une fin, et un accord ne vaut que pour un usage annoncé à l'avance. 0,5 par ligne.

- d Un élève quitte le lycée et veut que l'application supprime ses données ; une autre refuse que son numéro serve à lui envoyer de la publicité. Quel droit chacun exerce-t-il ? (1 pt)

✓ **Le premier exerce son droit à l'effacement ; la seconde, son droit d'opposition.**

- **Pourquoi.** L'effacement fait supprimer les données ; l'opposition permet de refuser un usage particulier, ici la publicité. 0,5 par droit.

⚠ CE QUI COÛTE DES POINTS

Un principe nommé sans la ligne qui le méconnaît : 0,5 sur 1.

Barème : 4 points — 1 par question.

🔍 ERREURS FRÉQUENTES

- Ne voir de donnée personnelle que dans le nom et le prénom.
- Citer un principe sans montrer, ligne à l'appui, en quoi le formulaire le méconnaît.

RAPPEL DE COURS



On parle de big data quand le volume, la vitesse d'arrivée et la variété des données dépassent ce qu'un traitement classique absorbe. On y cherche des corrélations sans hypothèse préalable : une corrélation n'est pas une cause.

- a Associe chacun des trois aspects du big data à un élément de la situation : le volume, la vitesse, la variété.

1,5 pt

✓ **Volume : des milliards de valeurs par jour. Vitesse : une mesure chaque seconde. Variété : des nombres, des images et des messages écrits.**

→ **Pourquoi.** Les trois aspects ne se confondent pas : la quantité totale, le rythme d'arrivée, la diversité des natures de données. 0,5 par association juste.

- b Sans hypothèse de départ, les analystes constatent que les jours où l'on vend le plus de parapluies, les bus ont le plus de retard. Peut-on dire que les parapluies retardent les bus ? Propose une autre explication.

1 pt

✓ **Non : c'est une corrélation, pas une cause. Une autre explication : la pluie fait à la fois vendre des parapluies et ralentir la circulation.**

→ **Pourquoi.** Deux grandeurs peuvent varier ensemble parce qu'une même cause agit sur les deux ; une corrélation trouvée sans hypothèse doit être expliquée, pas prise pour une cause. 0,5 pour le refus, 0,5 pour l'explication.

⚠ CE QUI COÛTE DES POINTS

La corrélation prise pour une cause : 0 sur 1.

Barème : 2,5 points — 1,5 + 1.

🔍 ERREURS FRÉQUENTES

- Attribuer la mesure envoyée chaque seconde au volume plutôt qu'à la vitesse.
- Admettre que les parapluies retardent les bus parce que les deux varient ensemble.



Barème de notation

Ce que le correcteur regarde, exercice par exercice.

20 pts

CE QUI EST ATTENDU	EXERCICE	POINTS	CE QUI FAIT PERDRE LE POINT
Le fichier vérifié avant tout calcul	1	4 points — 1 par question	Une moyenne qui compte la case vide pour zéro : 0,5 sur 1.
Les requêtes enchaînées, opérations nommées	2	6 points — 1 + 1 + 1 + 1,5 + 1,5	Un résultat juste sans les opérations qui y conduisent : la moitié des points de la question.
La structure imbriquée mise en table	3	3,5 points — 1 + 1 + 1,5	Une liste écrite dans une seule case du CSV.
Le formulaire analysé, lignes à l'appui	4	4 points — 1 par question	Un principe nommé sans la ligne qui le méconnaît : 0,5 sur 1.
Le big data reconnu, la corrélation interprétée	5	2,5 points — 1,5 + 1	La corrélation prise pour une cause : 0 sur 1.
Total		20 points	



Que faire de ta note ?

Le contrôle ne s'arrête pas au chiffre : voici la suite selon ton score.



Niveau 1 · Essentiel



Niveau 2 · Standard — tu es ici



Niveau 3 · Approfondi



16 à 20

NIVEAU SUPÉRIEUR

C'est le niveau attendu en classe, et tu l'as. Passe au niveau 3 — Approfondi : un croisement de fichiers à réaliser, une copie à corriger, une structure à concevoir et une décision à argumenter.



10 à 15

REFAIRE CE NIVEAU

Repère l'exercice où tu as perdu le plus de points et relis son rappel de cours. Refais le niveau 2 dans quelques jours ; vise 16 avant de passer au niveau 3.



moins de 10

NIVEAU INFÉRIEUR

Fais d'abord le niveau 1 — Essentiel : il vérifie le vocabulaire des fichiers et les opérations élémentaires dont ce contrôle a besoin. Reviens ensuite au niveau 2.

✓ COMPARE AVEC TA PRÉVISION

- Reprends la grille « Avant de rendre ta copie » : là où tu avais coché « acquis » et perdu des points, relis le rappel de cours de l'exercice.
- Refaire un contrôle quelques jours plus tard, sans regarder le corrigé, mesure ce que tu as retenu, et non ce dont tu te souviens.
- La fiche de cours du chapitre reprend chaque notion : ProgresSchool · SNT Seconde · Données — Structuration et traitement.