

L'étude de cas est ce que vous êtes en train de lire

Vingt jeux de données issus de dix-sept scripts à graine fixe qui se régénèrent à l'octet près, vingt fichiers de tests et 4 212 assertions, sept vérifications qui font échouer la compilation, et un schéma qui refuse les données non synthétiques. Chaque affirmation de ce cours se vérifie dans ce dépôt.

Pierre Robentz Cassion

Leçon 8 sur 8

Chaînes d'analyse reproductibles — Quelqu'un d'autre l'exécute

Ce que couvre cette leçon

- Pourquoi la plateforme est l'exemple
- Des données brutes produites, non collectées
- Une sortie commitée, pour une raison qui nomme sa condition
- Des vérifications à trois endroits, et la contrainte qui a décidé où
- La vérification qu'un graphe de références ne peut pas exprimer
- Ce que le dépôt refuse de contenir
- La reproductibilité au niveau du langage
- Ce que ce dépôt ne fait pas
- Rapportez-le en entier
- La suite

Pourquoi la plateforme est l'exemple

- Tous les autres cours d'ici emploient un jeu de données.

Des données brutes produites, non collectées — Shell

```
pnpm datasets:generate  
git status --short apps/data-analysis/public/datasets/files  
# (nothing)
```

Des données brutes produites, non collectées

- Vingt fichiers CSV, dix-sept scripts générateurs, et les régénérer tous ne change rien — Chaque générateur a une graine. . .
- C'est une propriété plus forte que « la chaîne se réexécute » — Cela signifie que réexécuter est une étape de. . .
- Et les défauts sont délibérés — Chaque entrée de la liste `knownIssues` d'un jeu de données correspond à un bloc de son. . .

Une sortie commitée, pour une raison qui nomme sa condition

- Parce que le déploiement tourne sur Cloudflare Pages sans TeX, sans Python et sans pandoc — La compilation copie des...
- C'est la règle générale de la leçon 1, appliquée — Commitez un artefact produit quand le reconstruire n'est pas à la...
- Et cela fonctionne parce que la sortie est déterministe — `SOURCE_DATE_EPOCH` fixe l'horodatage de pdfTeX et celui de...

Des vérifications à trois endroits, et la contrainte qui a décidé où

Où	Quoi	Pourquoi là
Schéma Zod	Types de champs, énumérations, champs requis	S'exécute sur une entrée, sans système de fichiers
<code>astro build</code>	Sept vérifications de niveau graphe	A besoin de toute la collection
<code>pnpm test</code>	Vérifications de système de fichiers	La compilation n'a pas <code>node:fs</code>

Des vérifications à trois endroits, et la contrainte qui a décidé où

- La compilation Astro pré-rend dans un worker Cloudflare sans `node:fs` — si bien qu'une vérification qui doit regarder...
- Placez la vérification là où elle peut s'exécuter — C'est une contrainte d'ingénierie réelle, et elle a produit un...

La vérification qu'un graphe de références ne peut pas exprimer — Exemple

REFERENCE_RULES verifies that a declared slug resolves.
It cannot verify that an entry is pointed at.

La vérification qu'un graphe de références ne peut pas exprimer

- Un cours pourrait donc être livré sans laboratoire ni exercice, et chaque barrière resterait verte — C'est pour cela. . .
- Toute règle du type « chaque X a au moins un Y » exige un test de cette forme — et reconnaître lesquelles de vos. . .

Ce que le dépôt refuse de contenir — Exemple

```
.refine((d) => d.dataQuality.synthetic === true, {  
  message: "Only synthetic or fully de-identified datasets may be published.",  
})
```

Ce que le dépôt refuse de contenir

- Un jeu de données non déclaré synthétique fait échouer la compilation — Ce public modélise des données de. . .
- Cela va de pair avec produire les données plutôt que les collecter — La garantie est structurelle aux deux bouts : le. . .

La reproductibilité au niveau du langage — Exemple

```
{ "packageManager": "pnpm@11.9.0", "engines": { "node": ">=22" } }
```

La reproductibilité au niveau du langage

- Le gestionnaire de paquets est figé, non seulement les paquets — Un pnpm différent résout le fichier de verrouillage. . .
- TypeScript est maintenu en 6.x avec la raison écrite — TypeScript 7 n'expose pas encore l'API programmatique dont. . .

Ce que ce dépôt ne fait pas

- Pas de conteneur — Un fichier de verrouillage plus un gestionnaire de paquets figé reproduit le côté JavaScript ; les. . .
- Pas de test de déterminisme de bout en bout en intégration continue — La régénération des jeux de données est. . .
- La documentation de passation est mince — `CLAUDE.md` et `docs/DEFINITION_OF_DONE.md` portent les décisions, ce qui est. . .
- Nommer les lacunes est l'essentiel — Une étude de cas qui revendique tout est une étude de cas dont personne ne peut. . .

Rapportez-le en entier — Exemple (suite)

Reproducibility of this platform

Data	20 CSV files generated by 17 seeded scripts. Regenerating all of them produces byte-identical output; <code>`pnpm datasets:generate`</code> followed by <code>`git status`</code> is the verification.
Checks	7 fail the build (schema, references, locale, translation parity, topic-sector, synthetic-only, programme spine). 20 test files and 4,212 assertions run first in CI.
Output	Course PDFs, slide decks, figures and notebooks are committed and deterministic. <code>SOURCE_DATE_EPOCH</code> pins pdfTeX and pandoc.
Environment	pnpm 11.9.0 and Node 22 pinned. TypeScript held at 6.x with the reason and the removal condition documented.

Rapportez-le en entier — Exemple (suite)

Refused	A dataset not declared synthetic fails the build.
Not done	No container; no CI job that regenerates everything and fails on a diff; handover documentation thinner than this course asks for.

- La dernière ligne est celle à copier dans votre propre version — Une déclaration de reproductibilité sans section « pas. . .

- Le fil qui traverse les vingt est le même — Un nombre dans un rapport de programme est une affirmation, et tout ce que...
- La dernière est celle qui rend les autres vérifiables — Une analyse que personne ne peut réexécuter est une analyse...

Lire la leçon complète, avec le code exécutable

`https://data-analysis.cassion.dev/fr/cours/reproducible-workflows/
repro-08-this-repository`