

# Mode d'emploi passage certifications développeur.se data

## **Sommaire**

[SECTION 1 - L'ÉPREUVE DE CERTIFICATION DU PARCOURS DÉVELOPPEUR·EUSE DATA](#)

[SECTION 2 - GRILLE D'ÉVALUATION INDIVIDUELLE](#)

[SECTION 3 - PROCÈS VERBAL JURY](#)

## SECTION 1 - L'ÉPREUVE DE CERTIFICATION DU PARCOURS DÉVELOPPEUR · EUSE DATA

### Mode d'emploi :

Cette section présente l'épreuve de certification du parcours développeur · se data dans sa globalité, elle donne des indications sur les modalités de passage (composition du jury, temporalités, etc.) ou encore sur les livrables attendus.

Cette section doit être partagée par le · a responsable de session :

- à l'ensemble de l'**équipe projet**, à minima **un mois avant le lancement de la promotion**
- aux **apprenants**, au plus tard **avant la fin de la prairie**
- aux **formateurs**, à minima **deux semaines avant le lancement de la promotion**
- aux **jury de certification**, au plus tard **deux semaines avant la session**

## 1. La composition du jury

La présentation du projet final se fait devant un jury composé d'au moins **deux professionnels** avec à minima **3 années d'expérience**, habilités à juger des réalisations et présentations des élèves : développeurs et chefs de projet compétents en programmation et gestion des données, formateurs data extérieurs à Simplon, etc.

**/!\ Le ou les formateurs ayant formé les apprenants ne peuvent pas faire partie du jury.**

## 2. Le projet chef d'oeuvre de l'apprenant · e ?

Le chef d'oeuvre est **un projet attestant de la maîtrise de l'ensemble des compétences visées** par la ou les certifications. il s'agit de réaliser une application - pendant la formation ou pendant leur stage/alternance en entreprise.

Si c'est en entreprise (ce qui est idéal), il faut obtenir préalablement l'accord avec l'entreprise d'accueil. Toutes les parties devront donc d'abord se mettre d'accord sur les compétences qu'ils pourront couvrir pour la réalisation du projet. Et pour la soutenance, sur la possibilité de présenter leur projet à l'extérieur (enjeux de confidentialité des données).

Il doit être **rédigé un rapport du projet chef d'oeuvre**. Ce rapport doit être remis auprès du/des formateurs et du jury avant la session d'examen.

Si l'on récapitule, l'apprenant doit :

1. Réaliser un projet chef d'oeuvre (cf. Cahier des charges du chef d'oeuvre).
2. Rédiger un rapport qui reprend les différentes étapes de conception et de production du projet.
3. Soumettre le rapport au jury.

Le rapport doit contenir :

- Entre 20 et 30 pages.
- Une introduction.
- Une première partie sur :
  - la compréhension besoin client.
  - Un état de l'art

- La traduction technique et choix technique du projet.
- Une seconde partie sur la mise en oeuvre du projet :
  - Gestion de projet.
  - Retours d'expérience sur les outils et techniques.
- Une troisième partie sur le bilan de projet et les améliorations.
- Une conclusion.
- Annexes possibles :
  - Cahier des charges rédigé par l'apprenant ou fourni par l'entreprise

### 3. Ce que doit contenir le projet

#### A. Conception et développement de la base de données

Sélectionner une base de données selon les enjeux du projet.

À partir de cela, normaliser un schéma de base de données répondant aux attentes du client : *utiliser une méthodes de conception tel que [Merise](#) par exemple.*

La base de données doit contenir au moins trois tables et couvrir les trois types de relations : *one to one, one to many, many to many.*

Puis en fonction du schéma, développer la base de données : *création des tables, relations, indexes, etc.*

Mettre en place des fichiers de sauvegardes : *CSV, XML, JSON, etc.*

À partir de cela, automatiser le processus de sauvegarde de la base données : *backup.*

Identifier les sources de données nécessaires à l'application : *API, Open Data, web scraping, etc.*

À partir de cette identification, collecter ces données.

Sauvegarder l'état des données à chaque étape du cycle de vie de la donnée : *brute, nettoyée, mise à jour, etc.*

## B. Alimenter la base de données

Préparer les données à l'import en base de données : *format, organisation, nettoyage, etc.*  
Créer un répertoire pour les métadonnées.

Alimenter la base de données en respectant la méthode d'import choisi : *script, ligne de commande, migration, etc.*  
Automatiser l'alimentation de la base de données, de manière à pouvoir importer des données en continue et automatiquement : *script, cron, stream, etc.*

Automatiser la recherche de doublons et d'erreurs dans les données stockées.  
À partir de cette recherche, supprimer et/ou mettre à jour les données en erreur.

## C. Optimisation de la base de données

Mesurer et rendre compte des performances des requêtes du projet.  
À partir de ces mesures, apporter au moins une mesure d'optimisation des requêtes : *optimisation des requêtes, indexation des données, choix du moteur de données, etc.*

## D. Exploiter la base de données

Définir la ou les requêtes de récupération des données en base de données du projet.  
À partir de cela, interroger la base de données selon la ou les requêtes établies.

Puis, appliquer un traitement aux données sélectionnées en accord avec les enjeux du client : *calculs, formattage, unité, etc.*  
Enfin, mettre à jour les données en base de données : *édition, correction, complément, etc.*

## E. Exposer et représenter les données visuellement

Choisir les formats de représentations visuelles des données : *tableau, graphique, 3D, etc.*

Puis, produire ces représentations visuelles à l'aide d'outil de visualisation de données.

Selon les cas, générer de nouvelles données à partir de celles existantes à l'aide de calculs, manipulations ou algorithme.

Choisir les supports de diffusion des représentations visuelles des données : *interface web, document, etc.*

Puis, établir une mise en conformité du processus de diffusion des données avec la législation relative aux données : *RGPD, droit d'auteur, etc.*

À partir de cela, réaliser les supports de diffusion : *document, mail, site web, etc.*

Enfin, mettre en place les règles d'accès aux supports de visualisation des données en accord avec les attentes du client : *gestion utilisateurs, mot de passe, etc.*

## F. Gestion de projet et qualité

Réaliser un planning prévisionnel et le faire valider par le client.

Présenter un historique des sources utilisées dans la recherche de solutions : *veille, stackoverflow, github issues, etc.*

Expliciter la démarche de recherche et la pertinence des sources.

## 4. Oral du candidat(e)

La présentation orale dure au **total 50 minutes** et se découpe de la manière suivante :

- Présentation du projet **20mn.**
- Présentation de la démo **10mn.**
- Echange avec le jury **20mn.**

Pendant la présentation orale, il faut veiller à ce que le propos soit clair et structuré, les arguments s'articulent de façon cohérente.

### A. La présentation du projet

La présentation à l'oral, appuyée d'un support (slides ou autre) devra comporter les éléments suivants :

*A noter : lorsque "screenshot" est mentionné ci-dessous, cela ne s'applique que aux apprenants qui présentent un projet d'entreprise et qui serait contraint d'anonymiser les données.*

*Dans les cas où la sécurisation des données n'est pas un enjeu, les parties "screenshots" pourront être présentées directement lors de la démonstration.*

### **Introduction du projet**

Un résumé de la demande exprimé par le client.

- Le contexte.
- la formulation de la problématique.
- La solution souhaitée/envisagée.

Résumé de la réponse finale apportée.

### **Analyse de la demande**

Les enjeux réglementaires pour le traitement des données de l'application.

Les utilisateurs du projet.

Le schéma fonctionnel utilisé pour le projet.

Un planning prévisionnel. Une présentation du macro-planning initial et réel cohérent au vu de la complexité du projet.

### **Conception technique**

Les choix technologiques dont les arguments justifiant de la technologie choisie pour réaliser la base de donnée.

Le schéma de la base de donnée s'appuyant sur une méthode de conception explicite.

### **Développement de la base de données**

Une présentation du processus d'alimentation de la base de donnée et les raisons de ce choix (origine des données, pré-traitement, import...).

Une présentation des méthodes de nettoyage de la base de donnée.

Screenshots :

- des fichiers de sauvegarde des données.
- des fichiers à intégrer dans la base de donnée.
- du répertoire pour les métadonnées.
- des fichiers de données sauvegardées.

Une présentation du processus d'entretien de la base de donnée : planning et acteurs.

---

Une présentation du protocole de recherche d'erreurs (automatisé et non-automatisé).

### **Optimisation de la base de données**

Une présentation des méthodes d'analyse des performances des requêtes.

Une présentation des mesures d'optimisation prises (indexation, forme normale, moteur de données...).

### **Exploitation de la base de données**

Une présentation du processus de sélection et traitement des données.

Une présentation du processus de mise à jour des données.

Un présentation du processus de test et de mise en production.

### **Visualisation des données**

Une esquisse des visualisations de données qui respecte les contraintes techniques et réglementaires ainsi que la demande du commanditaire.

Screenshot des visualisations de données élaborées.

### **Pour aller plus loin...**

Une présentation des points à améliorer pour le prochain projet.

Le résumé d'un sujet de veille réalisé à l'occasion du projet.

Le résumé d'un sujet de veille réalisé à l'occasion du projet et ayant permis une mise à jour sur les technologies utilisées.

Une bibliographies de liens et de ressources pour assurer une mise à jour de ses connaissances.

## **B. La Démonstration**

La démonstration doit présenter les éléments suivant :

### **Développement de la base de données**

Les sources des données utilisées.

La base de données SQL ou NoSQL. Il est possible d'alimenter cette base de données (insertion, suppression).

Un processus de sauvegarde automatisé.

Un dispositif pour autocontrôler, tout au long du processus de développement, la cohérence des données et la conformité à la demande.



### **Exploitation de la base de données**

La récupération des données en base et le traitement appliqué aux données (sélection, modification).

L'exportation des données en base selon les besoins liés au projet.

Une évaluation de la performance des techniques d'interrogation de la base de données.

Un protocole de gestion des droits d'accès/modification par utilisateur

### **Visualisation des données**

Une visualisation de ses données en utilisant des outils adaptés :

- Librairie de fonctions d'analyse statistique et de visualisation.
- Logiciels.

Si besoin, le projet permet d'accepter des données complémentaires pour faciliter la réalisation des supports au moyen de :

- Calculs.
- Manipulations.
- Algorithmes.

Le projet comprend des rendus visuel à la disposition des personnes destinataires.

Le projet dispose de plusieurs format de présentation des résultats selon le public visé.

## **C. Les questions du jury**

Le but de cet échange est de permettre au jury d'éclaircir, si besoin, certains points du projet présenté mais surtout de s'assurer que toutes les compétences sont bien acquises.

Exemple de questions :

### **Analyse de la demande**

Le schéma fonctionnel présenté ne comprend-il l'expression de besoin en SI ? Les interfaces ? Ou encore les protocoles de communication ? Si oui/non,

quel(le)s sont-ils/elles ?

Dans l'analyse de la demande, les utilisateurs sont-ils identifiés ? La nature des données ? Le problème est-il énoncé ?

### **Conception de la base de données**

Quels sont les trois types de relations en base de données (one to one, one to many, many to many) ? Comment les mettre en place ? Comment les exploiter ?

Quel(s) outil(s)/méthode(s) de conception a été utilisé pour concevoir la base de données ? Quel retour d'expérience sur cet/ces outil(s) ?

Une norme (forme normale, par exemple) a-t-elle été utilisée pour concevoir la base de données ? Laquelle ? Pourquoi ?

Comment les choix pour nettoyer les données brut ont-ils été pris ? Quels critères ? Pourquoi ?

Quels sont les formats de fichiers de données courants ? Pourquoi avoir fait le choix d'un tel type de format de fichier ?

Comment a été créée la structure de la base de données (tables, relations, etc.) ? Par quel moyen/outil/méthode (sql, interface, migrations, etc.) ?

Comment l'intégrité des données est-elle garantie ? Quels moyens sont-ils mis en oeuvre (triggers, tâches planifiées, etc.) ?

Comment les données sont-elles indexées ? Pourquoi avoir fait le choix de tel ou tel index ?

Le projet prévoit-il un système de backup ? Comment est-il mis en place ?

### **Exploitation de la base de données**

Comment les performances d'exploitation de la base de données sont-elles mesurées, quantifiées ? Les outils, méthodes ?

Quels sont les exports possibles ? Leur format ?

Comment les sélections de données en base impliquant des relations sont-elles réalisées ?

### **Visualisation des données**

Avez-vous maquetter, schématiser les visualisations avant de les produire ? Comment ?

Quels sont les règles et contraintes d'accès aux données et visualisation des données ? Y a-t-il un système de droit, mot de passe, utilisateurs ?

### **Pour aller plus loin...**

Quels sont les points d'améliorations identifiés ? Quelle priorité ?

Quelles ont été les difficultés rencontrées ? Comment avez-vous surmonté ces difficultés ? Quel moyens ? Quelles ressources ?