

**IUT de Vélizy-Rambouillet****CAMPUS DE VÉLIZY-VILLACOUBLAY***Durand Antonin**Joula Maxime**Parciany Benjamin**Zehren William*

Recueil des exigences

Table des matières

1. Chapitre 1 - Objectif et portée
 1. Quels sont la portée et les objectifs généraux ?
 2. Les intervenants. (Qui est concerné ?)
 3. Qu'est-ce qui entre dans cette portée ? Qu'est-ce qui est en dehors ? (Les limites du système.)
2. Chapitre 2 – Terminologie employée / Glossaire
3. Chapitre 3 – Les cas d'utilisation
 1. Les acteurs principaux et leurs objectifs généraux.
 2. Les cas d'utilisation métier (concepts opérationnels).
 3. Les cas d'utilisation système.
4. Chapitre 4 – La technologie employée
 1. Quelles sont les exigences technologiques pour ce système ?
 2. Avec quels systèmes ce système s'interfacera et avec quelles exigences ?
5. Chapitre 5 – Autres exigences
 1. Processus de développement
 1. Qui sont les participants au projet ?
 2. Quelles valeurs devront être privilégiées ? (exemple : simplicité, disponibilité, rapidité, souplesse etc...)
 3. Quels retours ou quelle visibilité sur le projet les utilisateurs et commanditaires souhaitent-ils ?
 4. Que peut-on acheter ? Que doit-on construire ? Qui sont nos concurrents ?
 5. Quels sont les autres exigences du processus ? (exemple : tests, installation, etc...)
 6. À quelle dépendance le projet est-il soumis ?
 2. Règles métier
 3. Performances
 4. Opérations, sécurité, documentation

5. Utilisation et utilisabilité
6. Maintenance et portabilité
7. Questions non résolues ou reportées à plus tard
6. Chapitre 6 – Recours humain, questions juridiques, politiques, organisationnelles.
 1. Quel est le recours humain au fonctionnement du système ?
 2. Quelles sont les exigences juridiques et politiques ?
 3. Quelles sont les conséquences humaines de la réalisation du système ?
 4. Quels sont les besoins en formation ?
 5. Quelles sont les hypothèses et les dépendances affectant l'environnement humain ?

Chapitre 1 - Objectif et portée

Quels sont la portée et les objectifs généraux ?

La portée ou le domaine du problème est une application web. Les objectifs généraux sont la simulation de calculs distribués dans différents domaines par trois profils d'utilisateur différents.

Les intervenants. (Qui est concerné ?)

Les intervenants sont les trois utilisateurs de l'application, à savoir, le visiteur ou l'utilisateur ne possédant pas de compte, l'utilisateur possédant un compte et enfin, l'administrateur.

Qu'est-ce qui entre dans cette portée ? Qu'est-ce qui est en dehors ? (Les limites du système.)

Ce qui rentre dans la portée est la réalisation des différents modules de simulations de calculs distribués, la mise en place de trois profils d'utilisation, le développement de différentes pages web, la conception d'une ou plusieurs bases de données. Il est également nécessaire de mettre en place l'installation et la sécurisation du serveur web, de la base de données et du cluster de raspberrypi.

Chapitre 2 – Terminologie employée / Glossaire

- **Abstraction** : Action intellectuelle consistant à mettre en évidence un élément en portant l'attention sur lui et en négligeant tout autre aspect. Elle fait partie des principes fondamentaux de l'approche orientée objet.
- **Calcul distribué** : Un calcul distribué, ou réparti ou encore partagé, est un calcul ou un traitement réparti sur plusieurs microprocesseurs et plus généralement sur plusieurs unités centrales informatiques. Le calcul distribué est souvent réalisé sur des clusters de calcul spécialisés, mais peut aussi être réalisé sur des stations informatiques individuelles à plusieurs cœurs. La distribution d'un calcul est un domaine de recherche des sciences mathématiques et informatiques. Elle implique notamment la notion de calcul parallèle.
- **Cas d'utilisation** : Exigences fonctionnelles.
- **Conception** : Activité du génie logiciel ayant pour objectif de fournir une structure pour un élément complexe. Lors de la conception on décompose l'élément en parties, on attribue des responsabilités à chaque partie et on s'assure que les parties sont bien assemblées afin d'accomplir une tâche bien définie.
- **Développement** : Activité du génie logiciel consistant à programmer le logiciel dans un langage de programmation.
- **Génie logiciel** : Discipline liée à tous les aspects de la production du logiciel complexe et avec d'importantes contraintes de qualité. Elle est liée à l'application de théories, de méthodes et l'utilisation d'outils pour le développement logiciel d'une façon professionnelle. Elle favorise et permet le travail en équipe.
- **Parallélisme** : En informatique, le parallélisme consiste à mettre en œuvre des architectures d'électronique numérique permettant de traiter des informations de manière simultanée, ainsi que les algorithmes spécialisés pour celles-ci. Ces techniques ont pour but de réaliser le plus grand nombre d'opérations en un temps le plus petit possible
- **Spécification** : Activité du génie logiciel consistant à décrire ce que le logiciel doit faire.
- **Haute disponibilité** : Objectif pour une application, service, site web d'être tout le temps disponible et fiable. Cette haute disponibilité peut être définir par deux fonctionnalités. La première est le **hors-panne** où une machine est remplacée par une autre dans le cas d'un dysfonctionnement. La deuxième fonctionnalité est **l'équilibrage de charge** où un service est exécuté par une machine en fonction de la charge des machines.

Chapitre 3 – Les cas d'utilisation

Les acteurs principaux et leurs objectifs généraux.

Les acteurs principaux sont les utilisateurs possédant un compte. Leurs objectifs généraux sont d'utiliser les différents modules de l'application afin de réaliser des calculs distribués.

Les cas d'utilisation métier (concepts opérationnels).

- L'utilisateur s'inscrit
- L'utilisateur se connecte
- L'utilisateur modifie ses informations personnelles comme son login, adresse mail, nom, prénom et mot de passe
- L'utilisateur se déconnecte
- L'utilisateur supprime son compte
- Le visiteur non connecté accède au site

Les cas d'utilisation système.

- Le système enregistre les login et les mots de passe des utilisateurs
- Le système enregistre le login, le mot de passe, la date et l'adresse ip lors d'une connexion échouée
- La page principale donne accès à tous les modules du site

Chapitre 4 – La technologie employée

Quelles sont les exigences technologiques pour ce système ?

L'application est hébergée sur un kit Cluster Hat. Un kit cluster hat est composé d'un raspberry pi 4 qui est le rpi principal du cluster. Le rpi principal est relié à un contrôleur qui permet de gérer à l'aide du port GPIO du rpi principal (Général Purpose Input/Output) jusqu'à 4 raspberry pi zero, qui sont les noeuds du cluster. Le rpi principal possède quatre ports USB 2.0, un port ethernet, un port micro sd, un port hdmi, un port micro hdmi ainsi qu'un port micro usb pour le relier au contrôleur. Le contrôleur quant à lui possède en plus du port pour la liaison au rpi principal, quatre autres ports micro usb pour connecter 4 rpi zero. Chaque rpi zero possède deux ports micro usb, un port micro sd, un mini port hdmi ainsi qu'un connecteur 40 pin compatible avec le cluster.

Le rpi principal est un **Raspberry Pi 4 Model B**. Le processeur, d'architecture *ARMv7* 32 bits, est composé de 4 coeurs chacun cadencé à 1.5Ghz. La taille de la mémoire ram est de 4Go.

Pour ce qui est des 4 autres rpi, ce sont des **Raspberry Pi Zero W Rev 1.1**, le processeur utilise l'architecture *ARMv6-compatible processor rev7 (v61)*. L'architecture est donc de type 32 bits. Le processeur possède 1 coeur cadencé à 1Ghz. La taille de la mémoire ram est de 512Mo.

Avec quels systèmes ce système s'interfacera et avec quelles exigences ?

Les différents noeuds du kit Cluster Hat doivent pour communiquer en ensemble ou communiquer avec le rpi host. Les différents sous-systèmes du kit Cluster Hat doivent être opérationnels comme le système ssh, et doivent de plus pouvoir communiquer entre eux comme les services MySQL et Web (comportant le serveur Apache avec PHP et nodeJs) de l'application.

Chapitre 5 – Autres exigences

Processus de développement

Qui sont les participants au projet ?

Les développeurs du projet:

- Antonin Durand
- Benjamin Parciany
- Maxime Jouglà
- William Zehren
- Thomas Dufaud

Le(s) client(s) du projet:

- Fabrice Hoguin

Quelles valeurs devront être privilégiées ? (exemple : simplicité, disponibilité, rapidité, souplesse etc...)

Le site web doit satisfaire les valeurs de disponibilité, robustesse, modularité, maintenabilité et portabilité, qui sont présentes dans la norme **ISO/IEC 25010:12**.

Le site web doit être accessible depuis n'importe quelle machine connectée à internet et à n'importe quand. L'application doit être utile et efficace pour le domaine d'expertise du client, en utilisant des calculs distribués par exemple. De plus la navigation dans le site doit être rapide, efficace et claire pour une expérience utilisateur positive. Ensuite, l'application doit pouvoir être déployée sur différents environnements d'exécution (Distribution Linux, Windows). Enfin, le site web doit être durable dans le temps, maintenable sans trop de complication, ainsi que évolutif facilement.

Quels retours ou quelle visibilité sur le projet les utilisateurs et commanditaires souhaitent-ils ?

Une communication est mise en place avec notre client pour valider le cahier des charges, les différents livrables.

a) Squelette du site

- Date de début du sprint : **24/11/2023**
- Date de fin du sprint : **04/12/2023**

Le premier sprint correspond à l'initialisation de l'application et de ses services, ainsi que du Kit Cluster hat hébergeant l'application. Premièrement, nous installerons les services Apache, PHP, MySQL dans le kit Cluster. Deuxièmement, un système de connexion, inscription et gestion de compte sera mis en place dans l'application.

b) 1er module et fonctionnalités utilisateurs admin

- Date de début du sprint : **04/12/2023**
- Date de fin du sprint : **25/12/2023**

Le deuxième sprint correspond à la réalisation du premier module de calcul en fonction du cahier des charges donné par un de nos professeurs. L'objectif est de calculer les nombres premiers entre 1 et n en utilisant le calcul distribué sur les 4 pi zero.

Durant ce sprint, seront développées les différentes fonctionnalités de l'utilisateur administrateur de l'application. Il pourra visualiser les différents utilisateurs de cette dernière, les rechercher, mais pourra aussi supprimer des comptes utilisateurs et visualiser les différentes actions effectuées par les utilisateurs inscrits sur l'application.

c) 2ème module

- Date de début du sprint : **03/01/2024**
- Date de fin du sprint : **22/01/2024**

L'objectif de ce troisième livrable est de développer un autre module pour l'application. L'utilisateur pourra, dans ce module, calculer une valeur approchée de PI en utilisant des techniques probabilistes. Ce calcul devra être effectué de manière distribuée sur les différents rpi zero du kit Cluster Hat.

c) 3ème module

- Date de début du sprint : **à définir**
- Date de fin du sprint : **à définir**

Dans ce sprint, sera développé un troisième module pour l'utilisateur inscrit, utilisant le machine learning. Ce dernier pourra envoyer une image de voiture à l'application qui devra retourner le modèle de la voiture. Cette image sera passée à plusieurs modèles qui prédiront l'image de manière parallèle, pour une meilleure prédiction.

Que peut-on acheter ? Que doit-on construire ? Qui sont nos concurrents ?

Aucun produit ne doit être acheté pour pouvoir utiliser pleinement le site web, aucun élément doit être construit. Pour se connecter au site web, il suffit de se connecter à internet. Les concurrents sont les autres groupes de SAE.

Quelles sont les autres exigences du processus ? (exemple : tests, installation, etc...)

Installation et prise en main du kit Cluster Hat. Tests de connexion en ssh entre les rpi pour vérifier la bonne installation du Cluster Hat.

À quelle dépendance le projet est-il soumis ?

Il y a une contrainte de temps, la SAE se termine en janvier. Le projet doit également mettre en place du parallélisme afin de profiter du cluster pi. Il y a également des contraintes de temps pour chaque sprint.

Règles métier

Performances

Le stockage des données dans la base de données doit être performante, ainsi que les calculs distribués et opérations effectuées dans les modules.

Opérations, sécurité, documentation

Le Raspberry Cluster HAT contenant le système doit être sécurisé. Les fonctions et le code doivent être documentés.

Utilisation et utilisabilité

Pour pouvoir utiliser pleinement le site, il suffit de se connecter à internet et de s'enregistrer sur l'application.

Maintenance et portabilité

Régler les éventuels bugs présents dans l'application.

Avoir un suivi des erreurs et opérations non voulues dans l'application.

Questions non résolues ou reportées à plus tard

Chapitre 6 – Recours humain, questions juridiques, politiques, organisationnelles.

Quel est le recours humain au fonctionnement du système ?

L'administrateur système doit vérifier dans les fichiers de logs si des connexions intempestives ont eu lieu.

Quelles sont les exigences juridiques et politiques ?

Dans la fiche n°1 "Identifier les données à caractère personnel", la pseudonymisation des données à caractère personnel est une mesure qui s'applique pour notre projet. En effet, l'utilisateur doit se connecter au site avec un login et un mot de passe, pour ensuite utiliser des modules. Le login ou le pseudonyme étant considéré selon la CNIL comme une donnée à caractère personnelle et est la seule dans le site que nous devons développer, ainsi cette dernière doit rester intègre et réduisant les risques pour les personnes concernées. Pour ce faire, le login va être crypté selon une certaine méthode.

Ensuite, dans la partie "Outils et pratiques" de la fiche n°2, nous avons effectué des choix technologiques. En effet, les systèmes installés sur le Raspberry Pi 4 tels que MariaDB, PHP, Apache ou encore le système d'exploitation Raspberry PI OS sont mis à jour pour obtenir les dernières versions de sécurité et éviter certaines vulnérabilités.

En ce qui concerne la gestion du code source dans la fiche n°4, nous avons paramétré le gestionnaire de code source GitHub pour assurer la sécurité du code, des données. Pour ce faire, une "authentification forte par clé SSH" a été mise en place, ce qui permet d'assurer une connexion sécurisée entre un ordinateur et GitHub.

La sécurité et l'accès aux données d'un serveur et d'un site web est une thématique très importante abordée dans la fiche n°6. Dans un premier temps, plusieurs mesures de sécurités ont été mises en place concernant le Raspberry Cluster HAT pour "sécuriser les infrastructures". Tout d'abord, comme indiqué précédemment, le système est régulièrement mis à jour, de plus, le seul moyen de se connecter au dernier est à distance via tunnel ssh, avec l'utilisateur "dev". Une tentative élevée de connexion successive bloque pendant une certaine durée la connexion entre la machine tentant de se connecter, ce qui diminue le risque de DDOS. Cela est permis grâce au protocole fail2ban.

Les mots de passes des utilisateurs "dev" et "root" ont aussi été changés pour des raisons de sécurité et une restriction "aux ports de diagnostic et de configuration" a aussi été mise en place.

Enfin, pour se connecter au SGBD mariaDB, seuls 2 utilisateurs sont possibles, à savoir "root" et "user". Ces deux derniers ont un mot de passe, ce qui permet d'éviter toute connexion nous voulues à la base de données, de plus l'utilisateur user ne possède qu'un droit de sélection, d'ajout et de suppressions dans les tables de la base blitzcalc, ce qui permet d'empêcher "la modification de la structure de la base de données".

Dans un second temps, des mesures de sécurité sont nécessaires côté site web, en limitant "la divulgation d'informations sur les comptes existants", ce qui permet limiter le risque qu'un compte soit piraté. Pour ce faire, nous avons fait en sorte de "généraliser les messages d'erreurs d'authentification" en affichant seulement la phrase "login/mdp incorrect" en cas de connexion échouée.

La version d'Apache2 et de PHP n'est pas visible. Mais aussi des mesures sont prises pour éviter l'injection de code dans les formulaires d'inscription, de connexion par exemple avec la fonction en PHP "htmlspecialchars". La taille minimale d'un mot de passe est de 8 caractères.

Dans la fiche n°7, il est important d'étudier quelles données sont nécessaires au traitement mises en œuvre, ainsi il faut minimiser les données collectées. Dans notre cas, seules trois données non sensibles pour l'utilisateur connecté sont stockées et utilisées, à savoir un login ou pseudonyme, un mot de passe.

Aucune autre information n'est stockée comme le nom, prénom, date de naissance, géolocalisation par exemple car ces données ne sont pas nécessaires pour le traitement.

Dans la fiche n°8, il est sujet de la gestion des utilisateurs, c'est-à-dire l'étude des différents droits que possède chaque personne. Pour ce faire, comme nous l'avons indiqué plus tôt, chaque personne a un identifiant unique pour se connecter au site web, ce qui permet de différencier ce qu'ils peuvent faire et ne peuvent pas faire. On peut prendre l'exemple de l'utilisateur "gestion" qui est l'admin du site web et peut supprimer un ou plusieurs utilisateurs, ou encore voir le nombre d'utilisations de chaque module, ce dernier ne peut pas accéder aux pages des modules, de profil d'un utilisateur normale. Et inversement, un utilisateur normal ne peut pas accéder à la page de l'admin pour des raisons de sécurité mais peut accéder aux pages des modules.

De plus, nous avons aussi mis en place un "système de journalisation" dans l'unique but de "tracer, détecter toutes anomalies liées à la sécurité" quand une connexion au site web avec un mauvais login/mdp. Dans un fichier de log est stocké seulement la date de la tentative, l'adresse IP, le login et le mot de passe tentés. Ensuite après 6 mois les données sont supprimés de la base de données.

Nous sommes également concernés par la fiche n°10 du guide RGPD : « Veiller à la qualité de votre code et sa documentation ». En effet, cette fiche concerne tous les projets informatiques y compris le nôtre. Afin de respecter ces recommandations, le code a été commenté tout au long de la programmation. La qualité du code a également été contrôlé, le code est correctement indenté, le nom des variables et des fonctions sont explicites et nous avons évité les redondances dans le code.

La fiche n°11 : « Tester vos applications » préconise d'effectuer des tests sur l'application en cours de développement afin de vérifier son bon fonctionnement. Comme pour la fiche précédente, cette mesure concerne tous les projets informatiques, le nôtre n'y échappe donc pas. Les tests que nous avons réalisés ont été intégrés au fur à mesure de l'avancer de la programmation de l'application. Ces tests ont été exécuté avec des jeux de données propres aux tests.

L'un des objectifs de l'application est de recueillir des données d'utilisation des modules du site afin de réaliser des séries statistiques, c'est pourquoi, notre projet est concerné par la fiche n°12 : « Informer les personnes ». Cette fiche indique qu'il faut être transparent sur l'utilisation des données personnelles des utilisateurs. De manière plus précise, il faut « que toute information ou communication relative au traitement de données à caractère personnel soit concise, transparente, compréhensible et aisément accessible ». Nous avons donc prévu d'implémenter un lien dans la page d'accueil redirigeant vers une page web explicitant toutes les informations que l'utilisateur doit savoir comme « l'identité et les coordonnées de l'organisme » ou les « finalités ». Un message d'information sera également mis en place lors de la création d'un compte sur le site afin de prévenir l'utilisateur du recueil de ses données.

Dans la continuité des données personnelles, la fiche n°13 : « Préparer l'exercice des droits des personnes » nous indique que les personnes dont les données sont traitées doivent avoir plusieurs droits sur celles-ci comme « droit d'accès, de rectification, d'opposition, d'effacement, à la portabilité et à la limitation du traitement ». Pour permettre l'exercice de ces droits, nous avons prévu de partager une adresse mail sur la page d'information du traitement des données personnelles des utilisateurs. Les utilisateurs pourront l'utiliser afin de nous contacter et jouir de leurs droits.

Notre site collecte des données personnelles dans les logs de connexion échouées, c'est pourquoi nous sommes concernés par la fiche n°14 : « Gérer la durée de conservation des données ». Les données nécessaires au bon fonctionnement du site sont des données de journalisation, nous avons donc choisi de supprimer les données d'échec de connexion 6 mois après leurs enregistrements.

La fiche n°15 : « Prendre en compte les bases légales dans l'implémentation technique » décrit les différentes bases légales existantes. Les bases légales étant « en quelque sorte la justification de l'existence même du traitement » des données personnelles, il est nécessaire de définir une base légale pour tout site traitant des données personnelles. La base légale de notre site est l'intérêt légitime car le traitement des données personnelles n'est pas susceptible d'affecter les droits et les libertés des personnes concernées.

Le seul cookie que nous utilisons stocke les deux facteurs de la multiplication du captcha. Nous ne sommes donc pas concernés par la fiche n°16 : « Analyser les pratiques en matière de traceurs sur vos sites et vos applications ».

Comme tous les sites informatiques, il est important de mettre en place des mesures de sécurité informatiques. C'est ce dont parle la fiche n°18 : « Se prémunir contre les attaques informatiques ». Notre site possède déjà des mesures contre les attaques informatiques comme la manipulation URL, le bourrage d'identifiants, l'injection de code indirecte et l'injection SQL. Nos mesures sont respectivement de « vérifier que l'émetteur de la requête est autorisé à accéder à la ressource associée », d'utiliser un captcha, « neutraliser les caractères utilisés pour l'insertion de script » et « utiliser des requête préparées ». Pour contrer les attaques par force brute et les programmes malveillants, nous prévoyons de sauvegarder régulièrement la base de données et d'imposer des mots de passe conformes aux recommandations en vigueur.

Quelles sont les conséquences humaines de la réalisation du système ?

Quels sont les besoins en formation ?

Aucune formation n'est requise pour pouvoir utiliser le site web.

Quelles sont les hypothèses et les dépendances affectant l'environnement humain ?