Cahier des Charges

Twyn BioManage Dashboard

1. Introduction

1.1. Contexte

1.1.1 Contexte Général

Le projet vise à développer une dashboard permettant la déduplication biométrique des images faciales. L'objectif est de s'assurer qu'un citoyen ne dispose pas de plusieurs enregistrements biométriques avec des identifiants différents afin d'éviter les erreurs et les fraudes. Ce système sera utilisé par une banque pour gérer des millions d'images de visages avec un temps de traitement aussi rapide que possible.

1.1.2 Contexte et Enjeux du Projet

Caixa est l'une des plus grandes banques du Brésil, avec un actif total d'environ 300,63 milliards de dollars. Le projet s'inscrit dans un contexte où la gestion des identifiants fiscaux, appelés CPF (Cadastro de Pessoas Físicas), est cruciale pour prévenir les fraudes et garantir l'intégrité des systèmes financiers et sociaux.

Le CPF (Cadastro de Pessoas Físicas)

Le CPF est l'identifiant fiscal des personnes physiques au Brésil. Il est équivalent au numéro de sécurité sociale (NSS) aux États-Unis ou au numéro d'identification fiscale (NIF) en France.

1.1.2.1. Caractéristiques du CPF

- C'est un numéro unique à 11 chiffres attribué par la Receita Federal do Brasil (l'administration fiscale brésilienne).
- Il est obligatoire pour toute transaction financière, comme ouvrir un compte bancaire, acheter un bien immobilier, ou déclarer des impôts.
- Les étrangers peuvent également obtenir un CPF s'ils ont des obligations fiscales ou veulent réaliser certaines démarches administratives au Brésil.

1.1.2.2. Types de Fraudes Liées au CPF

Au Brésil, certaines fraudes liées au CPF et aux prestations sociales, comme l'assurance chômage ou la retraite, impliquent différentes techniques illégales. Voici quelques-unes des fraudes les plus courantes :

1️⃣ Usurpation d'identité et CPF multiples

- Certains fraudeurs créent ou utilisent des CPF falsifiés (par exemple, un faux numéro ou un CPF appartenant à une personne décédée) pour percevoir plusieurs aides sociales sous différents noms.
- Ils peuvent également emprunter le CPF d'une autre personne (souvent avec son consentement) pour cumuler des prestations sociales qu'ils ne devraient pas recevoir.

2️⃣ Faux retraités ou "retirados fantasmas"

- Certains criminels continuent de percevoir la retraite d'un proche décédé en ne signalant pas son décès aux autorités.
- Ils peuvent également utiliser de faux documents médicaux pour obtenir une retraite anticipée pour invalidité.

3️⃣ Fraude aux allocations chômage

- Certains retraités se font passer pour chômeurs en utilisant de faux contrats de travail ou en simulant une perte d'emploi pour toucher l'assurance chômage tout en continuant à percevoir leur pension.
- Des employeurs complices peuvent déclarer un licenciement fictif, permettant ainsi à l'ex-employé de recevoir une indemnité chômage frauduleuse.

4️⃣ Complicités administratives et blanchiment de CPF

- Des fonctionnaires corrompus peuvent enregistrer de faux bénéficiaires dans les systèmes gouvernementaux pour détourner des fonds.
- Des criminels achètent des CPF d'individus en difficulté financière pour enregistrer des comptes fictifs et détourner des aides.

1.1.2.3. Mesures de Lutte Contre les Fraudes

Le Brésil a mis en place plusieurs mécanismes pour lutter contre ces fraudes :

- Croisement des bases de données : La Receita Federal et d'autres agences croisent les informations bancaires, fiscales et sociales pour détecter les incohérences.
- Reconnaissance biométrique : De plus en plus de services exigent une vérification faciale pour éviter les fraudes.
- Contrôles plus stricts : Les paiements sont liés aux registres de décès pour bloquer les pensions versées aux morts.

1.2. Objectifs

- Détecter et supprimer les doublons d'images biométriques dans une base de données.
- Comparer les images 1 à N en utilisant une API biométrique pour identifier les visages similaires.
- Conserver uniquement une image par IDN (Identifiant National).
- Automatiser le processus de traitement et de comparaison des images.
- Permettre une gestion efficace des conflits et des erreurs.

1.3. Etude de l'Existant

L'étude de l'existant permet de déterminer les points faibles et les points forts d'un produit actuel pour déterminer les besoins du client, en vue d'en prendre en considération lors de la conception et la réalisation de notre plateforme.

1. Face++ (Megvii)

- **Description** : Face++ est une plateforme de reconnaissance faciale développée par Megvii, une entreprise chinoise. Elle est largement utilisée pour la vérification d'identité, la surveillance et la gestion des accès.
- **Ce qu'il offre** :
    - Reconnaissance faciale en temps réel.
    - Vérification d'identité et déduplication biométrique.
    - Analyse des émotions, de l'âge et du genre.
- **Points Forts :**
    - Très haute précision dans la reconnaissance faciale.
    - Utilisé dans des applications grand public et industrielles.
    - API facile à intégrer.
- **Points Faibles :**
    - Préoccupations concernant la confidentialité des données.
    - Dépendance à l'égard de la connectivité Internet pour les services cloud.

2. SenseTime (Plateforme de Reconnaissance Faciale)

- **Description** : SenseTime est une entreprise chinoise spécialisée dans l'intelligence artificielle et la reconnaissance faciale. Leur plateforme est utilisée pour la surveillance, la sécurité et le marketing.
- **Ce qu'il offre :**
  - Reconnaissance faciale en temps réel.
  - Analyse des émotions, de l'âge et du genre.
  - Déduplication biométrique pour éviter les doublons.
- **Points Forts :**
  - Technologie de pointe en reconnaissance faciale.
  - Utilisé dans des projets de grande envergure en Chine.
  - Intégration avec des systèmes de surveillance.
- **Points Faibles :**
  - Préoccupations concernant la confidentialité et l'utilisation des données.
  - Complexité d'intégration pour les petites entreprises.

3. FaceFirst (Solution de Reconnaissance Faciale)

- **Description** : FaceFirst est une plateforme de reconnaissance faciale utilisée pour la sécurité, la gestion des accès et la vérification d'identité. Elle est utilisée dans les aéroports, les magasins et les stades.
- **Ce qu'il offre :**
  - Reconnaissance faciale en temps réel.
  - Vérification d'identité et déduplication biométrique.
  - Intégration avec des systèmes de sécurité existants.
- **Points Forts :**
  - Haute précision et rapidité.
  - Utilisé dans des environnements critiques comme les aéroports.
  - Facile à intégrer avec des systèmes de sécurité.
- **Points Faibles :**
  - Coût élevé pour les petites entreprises.
  - Nécessite une infrastructure robuste.

3. BioID (Solution de Reconnaissance Faciale et Déduplication Biométrique)

- **Description** : BioID est une plateforme de reconnaissance faciale qui offre des services de vérification d'identité et de déduplication biométrique. Elle est utilisée pour la sécurité, la gestion des accès et la prévention des fraudes.
- **Ce qu'il offre :**
  - Reconnaissance faciale en temps réel.

    - Comparaison 1:N pour la déduplication biométrique.
    - Vérification d'identité et analyse des émotions.
- **Points Forts :**
    - Haute précision dans la comparaison 1:N.
    - Utilisé pour la prévention des fraudes et la gestion des identités.
    - API facile à intégrer.
- **Points Faibles :**
    - Coût élevé pour les grandes bases de données.
    - Préoccupations concernant la confidentialité des données.

5. TrueFace (Solution de Reconnaissance Faciale)

- **Description** : TrueFace est une plateforme de reconnaissance faciale utilisée pour la vérification d'identité, la sécurité et la gestion des accès. Elle est utilisée dans les secteurs de la banque, de la sécurité et du retail.
- **Ce qu'il offre :**
    - Reconnaissance faciale en temps réel.
    - Vérification d'identité et déduplication biométrique.
    - Intégration avec des systèmes de sécurité existants.
- **Points Forts :**
    - Haute précision et rapidité.
    - Utilisé dans des environnements critiques comme les banques.
    - Facile à intégrer avec des systèmes de sécurité.
- **Points Faibles :**
    - Coût élevé pour les petites entreprises.
    - Nécessite une infrastructure robuste.

Critères de Comparaison

1. Précision : La capacité du système à identifier et à comparer les visages avec exactitude.
2. Scalabilité : La capacité du système à gérer de grandes bases de données et un volume élevé de demandes.
3. Intégration : La facilité avec laquelle le système peut être intégré à des applications et services existants.
4. Coût : Le coût total d'utilisation du système, y compris les licences et l'infrastructure.
5. Sécurité : Le niveau de sécurité et de conformité aux normes internationales.
6. Flexibilité : La capacité du système à s'adapter à différents cas d'utilisation et secteurs.

Tableau de Comparaison

Critères	Face++ (Megvii)	SenseTime	FaceFirst	BioID	TrueFace
Précision	Très haute précision.	Technologie de pointe, haute précision.	Haute précision et rapidité.	Haute précision dans la comparaison 1:N.	Haute précision et rapidité.
Scalabilité	Scalable pour de grandes bases de données.	Scalable pour des projets de grande envergure.	Scalable pour des environnements critiques.	Scalable pour de grandes bases de données.	Scalable pour des environnements critiques.
Intégration	API facile à intégrer.	Complexe, mais puissante.	Facile à intégrer avec des systèmes de sécurité.	API facile à intégrer.	Facile à intégrer avec des systèmes de sécurité.
Coût	Coût modéré, mais dépendant du volume.	Coût élevé pour les grandes bases de données.	Coût élevé pour les petites entreprises.	Coût élevé pour les grandes bases de données.	Coût élevé pour les petites entreprises.
Sécurité	Conforme aux normes, mais préoccupations de confidentialité.	Haute sécurité, mais préoccupations de confidentialité.	Haute sécurité, utilisé dans des environnements critiques.	Haute sécurité, conforme aux normes.	Haute sécurité, utilisé dans des environnements critiques.
Flexibilité	Flexible pour des applications grand public et industrielles.	Flexible pour des projets de grande envergure.	Flexible pour des environnements critiques.	Flexible pour la prévention des fraudes.	Flexible pour des environnements critiques.

Résumé de la Comparaison

- Précision : Tous les systèmes offrent une haute précision, mais Face++ et SenseTime se distinguent par leur technologie de pointe, tandis que BioID excelle dans la comparaison 1:N.
- Scalabilité : Face++, SenseTime, et BioID sont très scalables pour de grandes bases de données, tandis que FaceFirst et TrueFace sont adaptés aux environnements critiques.
- Intégration : Face++ et BioID sont plus faciles à intégrer grâce à leurs API, tandis que SenseTime nécessite une expertise technique.
- Coût : Tous les systèmes peuvent être coûteux, mais Face++ offre une option plus modérée pour les petites entreprises.
- Sécurité : Tous les systèmes sont conformes aux normes de sécurité, mais FaceFirst et TrueFace sont particulièrement adaptés aux environnements critiques.
- Flexibilité : BioID est flexible pour la prévention des fraudes, tandis que Face++ et SenseTime sont adaptés à des applications plus larges.

---

2. Besoins Fonctionnels

2.1. Gestion des processus de déduplication

- **préparer** des données
- **Démarrer** le processus de traitement.
- **Mettre en pause** le processus.
- **Reprendre** un processus en pause.
- **Arrêter et réinitialiser** un processus en cours.
- Afficher la liste des processus de déduplication avec :
    - L'utilisateur ayant lancé le processus.
    - Date et heure de début et de fin.
    - Nombre d'images traitées, erreurs détectées, conflits détectés et IDN en doublon.
- Notifications en temps réel pour les admin et le super admin en cas de conflits ou d'erreurs.

2.2. Création d'historique :

- enregistrer les processus de chaque admin dans un historique
- préciser les détails(date,etat) de chaque processus par chaque admin

- Ajouter des filtres pour rechercher l’historique par date, admin, ou statut du processus.
- Permettre l’exportation des rapports (e.g., "Export Déduplication Results").

2.3. Gestion des conflits

- Identifier les conflits lorsque plusieurs utilisateurs lancent simultanément des processus.
- Afficher les détails des conflits détectés.

2.4. Partie Statistique du Dashboard

- **Objectif** : Ajouter une section dédiée aux statistiques dans le dashboard, séparée des autres processus de gestion de déduplication.
- **Fonctionnalités** :

  Suivi des Processus : Afficher des statistiques en temps réel sur l'état des processus de déduplication en cours, tels que :

  - Nombre de processus démarrés.
  - Nombre de doublons détectés.
  - Nombre d’images traitées.
  - Temps moyen par processus.

  Analyse des Erreurs : Visualiser des statistiques liées aux erreurs et conflits survenus lors du traitement des images :

  - Nombre d'erreurs (API échouées, images corrompues, etc.).
  - Taux d'échec global.

  Performances : Indicateurs de performance du système :

  - Temps de traitement par image.

  Visualisations : Graphiques et tableaux pour une analyse facile des données :

  - Graphiques en barres, en camembert, ou en courbes pour suivre les métriques sur une période donnée.
  - Tableaux détaillant les processus avec un accès rapide aux détails de chaque processus.

  Alertes : Notifications en temps réel concernant des anomalies ou des performances faibles (par exemple, si un processus dépasse un certain temps de traitement ou génère un nombre élevé d'erreurs).

Ces statistiques seront essentielles pour que les administrateurs et super administrateurs suivent la performance du système et prennent des décisions basées sur les données.

---

3. Besoins Non Fonctionnels

- Performance : L'application doit être capable de traiter un grand volume d'images avec un temps de réponse rapide.
- Sécurité : Gestion des accès et authentification des utilisateurs.
- Scalabilité : Possibilité d'étendre l'application pour gérer un plus grand nombre de processus simultanés.
- Ergonomie : Interface utilisateur intuitive et facile d'utilisation.
- Fiabilité : Minimisation des erreurs lors du traitement et récupération en cas d'échec.

---

4. Diagrammes UML

4.1. Diagramme de Cas d'Utilisation

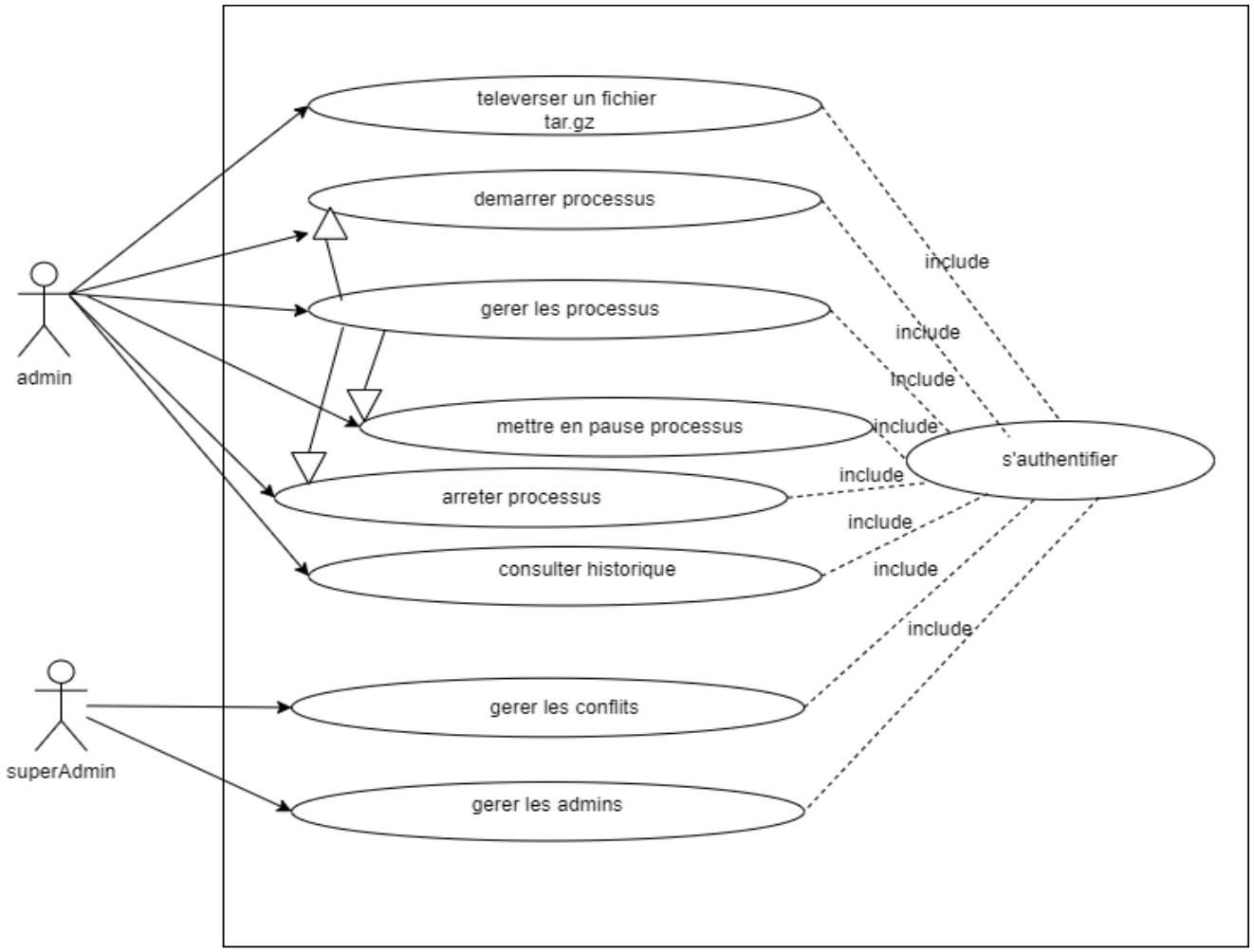
Acteurs :

1. **Admin :**
2. **SuperAdmin (hérite de Admin) :**

Cas d'Utilisation :

- **Admin :**
    - Téléverser un fichier .Tar d'images.
    - Démarrer un processus de déduplication.
    - Mettre en pause ou arrêter un processus.
    - Consulter l'historique des processus.

- **SuperAdmin :**
    - Gérer les conflits entre les processus.
    - Prioriser ou annuler des processus.



4.2. Diagramme de Séquence

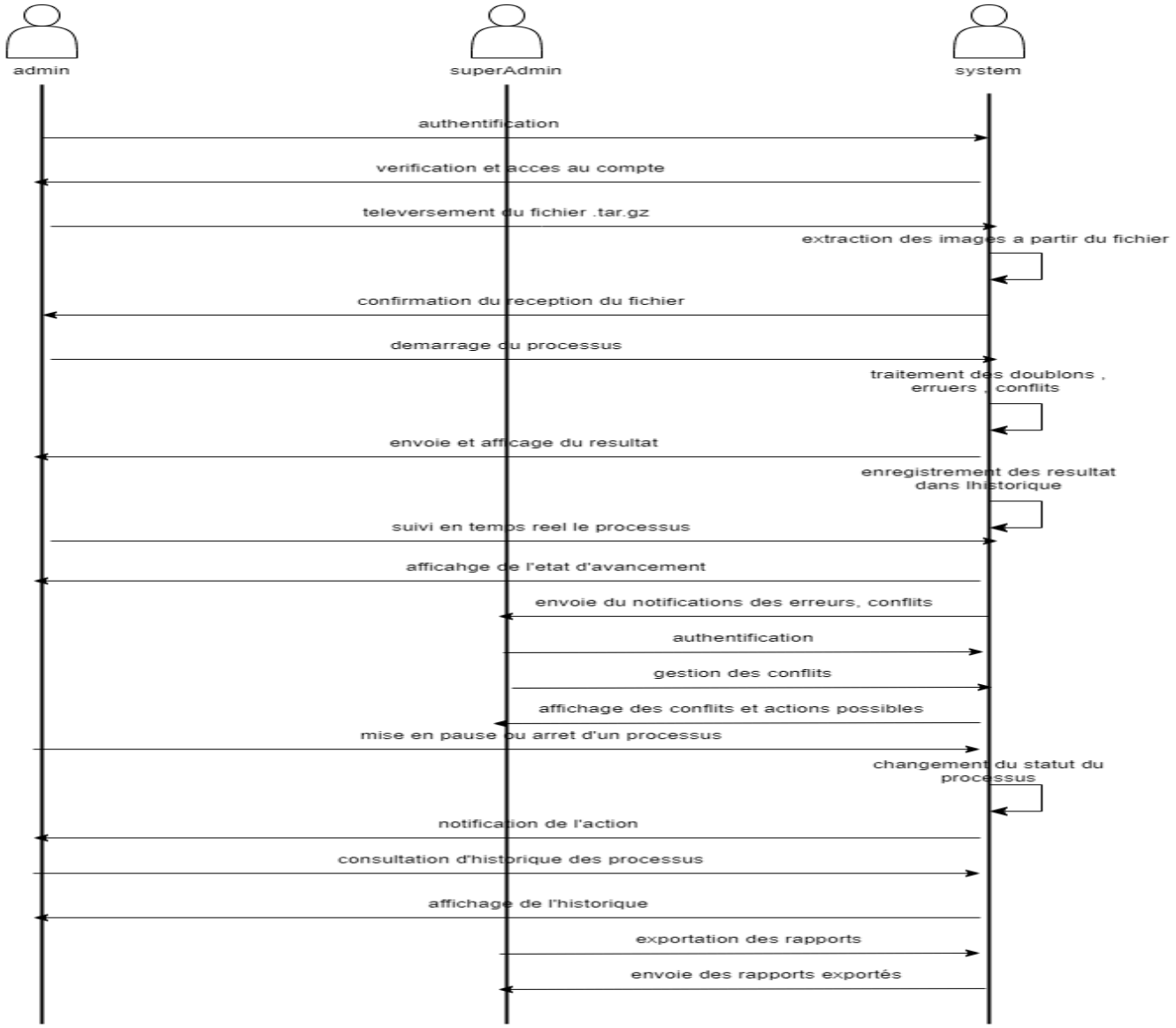
Acteurs : Admin, SuperAdmin, Système.

Séquences :

1. **Authentification :**
    - Admin, SuperAdmin, et Auditeur Système s'authentifient.
2. **Téléversement des Images :**
    - Admin : Téléverse un fichier .Tar contenant des images.
    - Système : Reçoit et extrait les images du fichier.Tar.
3. **Démarrage du Processus de Déduplication :**

- Admin : Lance le processus de déduplication.
- Système : Envoie les images extraites pour une comparaison.

4. **Comparaison des Images :**
   - Système : Envoie les images à l'API biométrique pour la comparaison des visages (1 à N).
   - API Biométrique : Effectue la comparaison des images pour détecter les doublons.
   - API Biométrique : Retourne les résultats au système.
5. **Affichage des Résultats :**
   - Système : Affiche les résultats de la déduplication (doublons détectés, erreurs, etc.).
6. **Enregistrement des Résultats :**
   - Système : Enregistre les résultats de la déduplication.
7. **Suivi du Processus en Temps Réel :**
   - Admin : Suivre l'avancement du processus en temps réel via une interface.
   - Système : Affiche l'état d'avancement du processus en temps réel.
   - Système : Envoie des notifications aux Admin et SuperAdmin.
8. **Gestion des Conflits :**
   - SuperAdmin : Gère les conflits qui peuvent survenir lors de l'exécution des processus.
   - SuperAdmin : Peut prioriser ou annuler certains processus en conflit.
9. **Gestion des Processus :**
   - Admin : Peut mettre en pause ou arrêter un processus en cours.
   - Système : Réagit en fonction de l'action de mise en pause ou d'arrêt.



4.3. Diagramme de Classe

Les Classes :

- **User**
  - **Attributs:**
    - ID : Identifiant unique.
    - Nom : Nom de l’admin.
    - Prénom : Prénom de l’admin.
    - Email : Adresse email.
    - MotDePasse : Mot de passe.
- **Admin Extends User:**

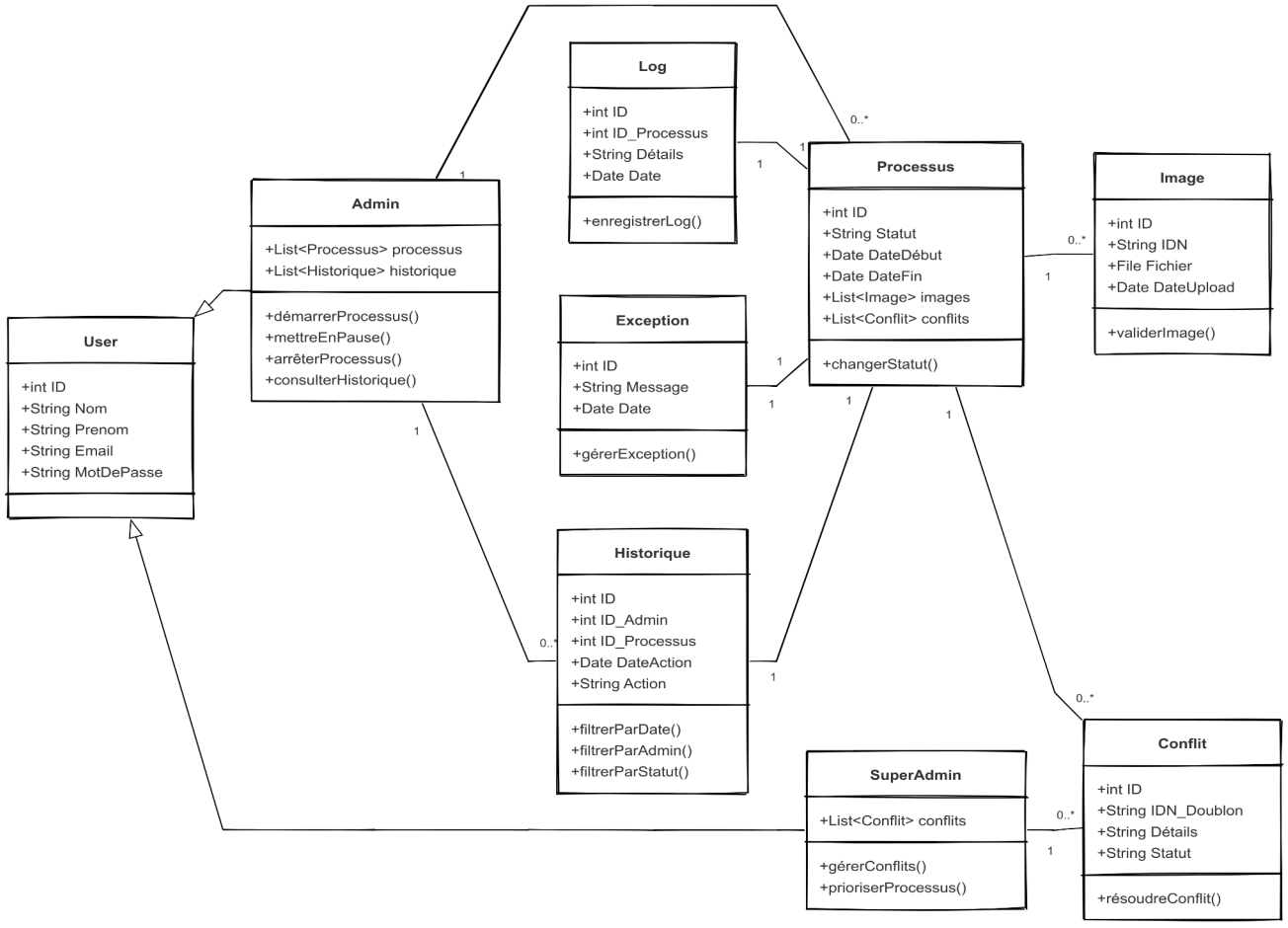
    - **Attributs :**
        - processus : Liste des processus démarrés.
        - historique : Liste des actions passées.
    - **Méthodes :**
        - démarrerProcessus() : Démarre un processus.
        - mettreEnPause() : Met en pause un processus.
        - arrêterProcessus() : Arrête un processus.
        - consulterHistorique() : Consulte l'historique des processus.
    - **Relations :**
        - Un Admin peut démarrer plusieurs Processus.
        - Un Admin peut consulter plusieurs enregistrements dans Historique.
- **SuperAdmin Extends User :**
    - **Attributs :**
        - conflits : Liste des conflits gérés.
    - **Méthodes :**
        - gérerConflits() : Gère les conflits entre les processus.
    - **Relations :**
        - Un SuperAdmin peut gérer plusieurs Conflits.
- **Image :**
    - **Attributs :**
        - ID : Identifiant unique.
        - IDN : Identifiant National de l'utilisateur.
        - Fichier : Fichier de l'image.
        - DateUpload : Date de téléversement.
    - **Méthodes :**
        - validerImage() : Valide l'image avant traitement.
    - **Relations :**
        - Une Image est traitée dans un Processus.
- **Processus :**
    - **Attributs :**
        - ID : Identifiant unique.
        - Statut : Statut du processus (En cours, Terminé, En pause).
        - DateDébut : Date de début du processus.
        - DateFin : Date de fin du processus.
        - images : Liste des images traitées.
        - conflits : Liste des conflits générés.
        - Priorité : Priorité du processus.
    - **Méthodes :**
        - changerStatut() : Change le statut du processus.
    - **Relations :**
        - Un Processus peut contenir plusieurs Images.

        - Un Processus peut générer plusieurs Conflits.
- **Conflit :**
    - **Attributs :**
        - ID : Identifiant unique.
        - IDN_Doublon : Identifiant National de l'utilisateur concerné.
        - Détails : Détails du conflit.
        - Statut : Statut du conflit (Résolu, Non résolu).
    - **Méthodes :**
        - résoudreConflit() : Résout le conflit.
    - **Relations :**
        - Un Conflit est associé à un Processus.
- **Historique :**
    - **Attributs :**
        - ID : Identifiant unique.
        - ID_Admin : Identifiant de l'admin ayant effectué l'action.
        - ID_Processus : Identifiant du processus concerné.
        - DateAction : Date et heure de l'action.
        - Action : Description de l'action.
    - **Méthodes :**
        - filtrerParDate() : Filtre l'historique par date.
        - filtrerParAdmin() : Filtre l'historique par admin.
        - filtrerParStatut() : Filtre l'historique par statut.
    - **Relations :**
        - Un Historique est associé à un Processus.
- **Log :**
    - **Attributs :**
        - ID : Identifiant unique.
        - ID_Processus : Identifiant du processus concerné.
        - Détails : Détails du log.
        - Date : Date et heure du log.
    - **Méthodes :**
        - enregistrerLog() : Enregistre un nouveau log.
    - **Relations :**
        - Un Log est associé à un Processus.
- **Exception :**
    - **Attributs :**
        - ID : Identifiant unique.
        - Message : Message d'erreur.
        - Date : Date et heure de l'exception.
    - **Relations :**
        - Une Exception est associée à un Processus.

Relations entre les Classes :

- Admin et Processus : Un Admin peut démarrer plusieurs Processus.
- Admin et Historique : Un Admin peut consulter plusieurs enregistrements dans Historique.
- SuperAdmin et Conflit : Un SuperAdmin gère plusieurs Conflits.
- Processus et Image : Un Processus contient plusieurs Images.
- Processus et Conflit : Un Processus peut générer plusieurs Conflits.
- Historique et Processus : Un Historique est lié à un Processus.
- Log et Processus : Un Log est associé à un Processus.
- Exception et Processus : Une Exception est associée à un Processus.

Voir le Diagramme :



---

5. Technologies

5.1 Frontend

- **Technologie** : **Angular 17**
    - Utilisation de la dernière version stable d'Angular pour une interface utilisateur réactive et performante.
    - Angular 17 permet d'implémenter une gestion efficace des processus en temps réel et une interface intuitive pour l'administration des tâches de déduplication.

5.2 Backend

- **Technologie** : **.NET Core Web API**
    - **.NET Core** est utilisé pour développer l'API web qui gère la logique de traitement des images et les interactions avec la base de données.
    - Cette API sera responsable de la gestion des processus de déduplication, de la comparaison des images via l'API biométrique et de la gestion des différents statuts des processus.

5.3 Base de Données

- **Technologie** : **RavenDB**
    - **Base de données principale** : Contient toutes les images faciales et les informations personnelles des utilisateurs, telles que les identifiants (IDN, CPF, etc.).
    - Elle est utilisée comme source pour la déduplication.
    - **Base de données locale** : Cette base stocke les images extraites et préparées pour le traitement, spécifiquement pour le processus de déduplication.
    - **Base de données Clean et Blacklist** : Après le processus de déduplication, les résultats sont enregistrés dans deux bases de données :
        - **Clean Database** : Contient les images validées sans doublons, prêtes pour une utilisation légale et fiable.
        - **Blacklist Database** : Contient les images suspectes ou frauduleuses, identifiées lors des comparaisons.
    - **Base de données Personne** : Contient l'ID d'une personne ainsi que son visage correct pour la référence lors de la déduplication.
    - **Base de données Visages Similaires** : Contient les visages similaires détectés, associés à leurs identifiants, pour permettre la gestion des doublons ou fraudes.
- *Note : RavenDB est un choix pertinent pour ce projet car il supporte une gestion rapide de grandes quantités de données non relationnelles, et il offre des fonctionnalités de recherche et de traitement de documents efficaces.*

5.4 CI/CD

- **Technologie** : **Azure DevOps**
    - Azure DevOps sera utilisé pour mettre en place un pipeline CI/CD (Intégration Continue / Déploiement Continu), permettant l'automatisation des tests unitaires, fonctionnels, et du déploiement de l'application.
    - **Automatisation des tests** : Assurer que chaque modification du code n'introduit pas de régressions.
    - **Mise à jour continue de l'application** : Déployer les nouvelles versions de l'application de manière fluide et sans interruption pour les utilisateurs.

5.5 Hébergement

- **Technologie** : **Azure DevOps** pour l'hébergement et le déploiement
    - L'application sera hébergée sur Azure, ce qui permettra une scalabilité, une sécurité et une performance optimales pour traiter des millions d'images faciales.
    - Azure DevOps intégrera à la fois les processus de développement et de déploiement, simplifiant la gestion de l'infrastructure et le suivi de la performance de l'application.

5.6 Compression et Téléchargement des Fichiers

- **Format de compression** : **tar.gz**
    - **tar.gz** sera utilisé pour compresser les fichiers d'images afin de faciliter leur téléchargement et leur traitement. Ce format est efficace pour les grandes quantités de données.
    - L'interface utilisateur permettra le téléversement de fichiers compressés au format **tar.gz**, que le système extraira et traitera pour la déduplication.

---

6. Conclusion

Le projet offre une solution innovante et robuste pour la gestion et la déduplication biométrique des images faciales. En réponse aux exigences de sécurité de la banque, il permet de garantir un accès sécurisé et unique aux utilisateurs grâce à l'authentification par reconnaissance faciale. Cette approche évite les risques de

fraude liés aux méthodes d'authentification traditionnelles, telles que les mots de passe ou les empreintes digitales.

En outre, l'intégration d'une API biométrique avancée permet une déduplication rapide et précise des images faciales, assurant ainsi que les doublons soient efficacement supprimés tout en maintenant une performance optimale dans la gestion des données. Le dashboard de gestion offre un contrôle administratif complet, facilitant la supervision des processus et la gestion des conflits, ce qui est crucial pour un environnement bancaire.

Ce projet répond non seulement aux besoins stricts de sécurité, mais aussi à ceux liés à la gestion des données dans un contexte où la confidentialité et l'intégrité sont primordiales. Le pipeline DevOps mis en place assure une automatisation efficace des déploiements, garantissant la mise à jour continue et sans erreur du système.

En somme, ***le projet*** contribue à renforcer la sécurité des utilisateurs tout en offrant une solution performante et scalable, parfaitement adaptée aux exigences du secteur bancaire. Grâce à cette solution biométrique, la banque peut offrir à ses clients un moyen sûr et rapide d'accéder à leurs services, tout en assurant la conformité aux normes de sécurité et de confidentialité.