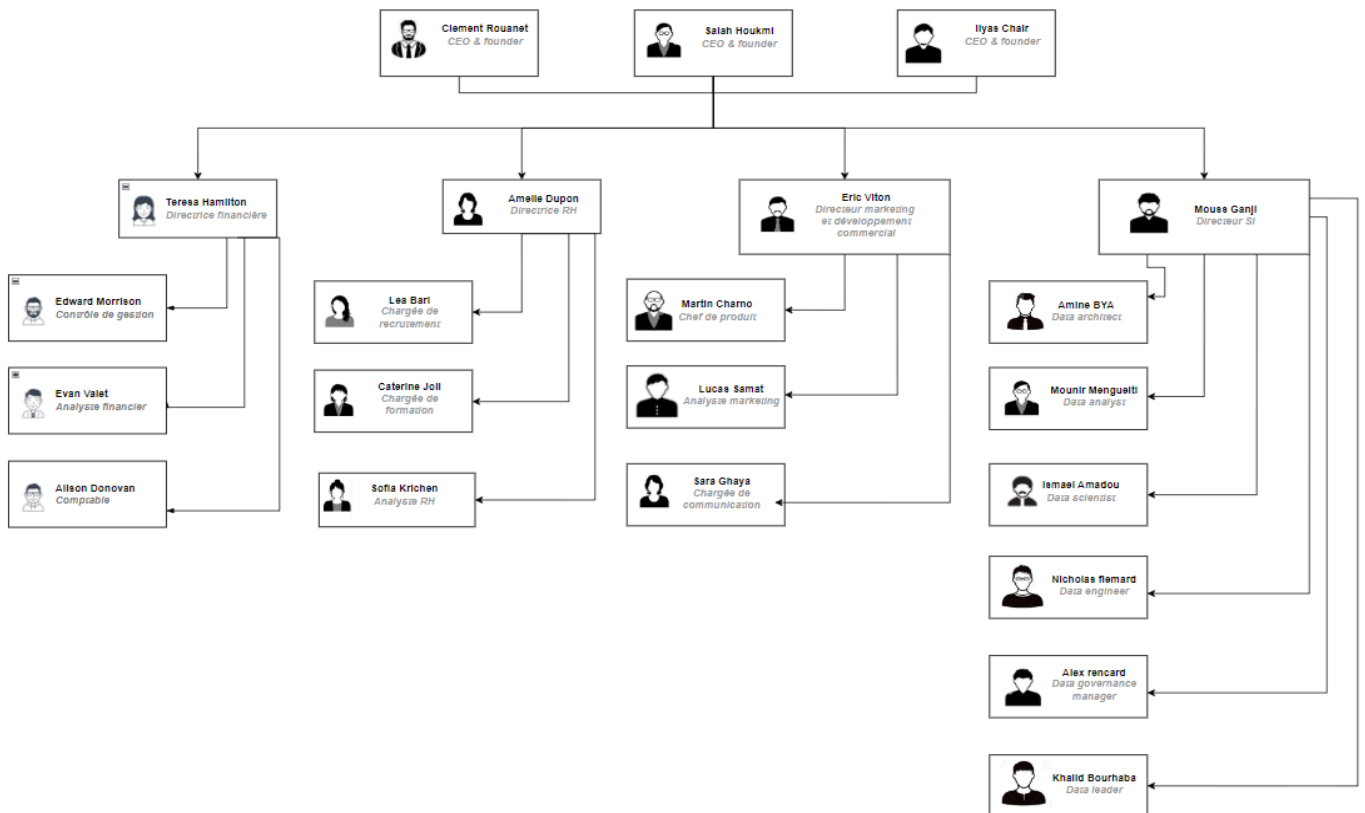


Projet final

Data Management & gouvernance

Question 1 :



Dans le contexte de la Banque Efrei, l'utilisation d'AWS Cloud comme plateforme principale pour la gestion des données constitue une approche moderne et évolutive. Voici comment l'architecture et les technologies utilisées par l'entreprise s'intègrent dans le processus de flux de données, en lien avec les types de plates-formes de données pertinents pour répondre aux objectifs stratégiques de l'entreprise :

1. AWS Cloud avec S3 Bucket :

- AWS S3 est utilisé comme un Data Lake pour stocker les données brutes provenant de diverses sources de l'entreprise, telles que les transactions financières, les données RH, les interactions clients, etc.
- Les données sont stockées de manière scalable, sécurisée et hautement disponible, permettant un accès rapide et flexible pour les traitements ultérieurs.

2. Lambda Functions et amazon Sagemaker:

- Les Lambdas sont des fonctions de calcul sans serveur utilisées pour le traitement des données en temps réel ou par lots.
- Les fonctions Lambda peuvent être utilisées pour automatiser des tâches telles que la validation des données, la normalisation et la transformation initiale.
- Les modèles de machine learning peuvent être déployés et exécutés via AWS Lambda pour effectuer des analyses avancées et fournir des prédictions basées sur les données.
- SageMaker permet aux équipes de choisir et d'expérimenter rapidement avec différentes approches de modélisation, en accélérant ainsi le processus de développement. De plus, avec ses fonctionnalités intégrées de suivi et de gestion des modèles, SageMaker simplifie la gestion du cycle de vie des modèles, depuis l'entraînement initial jusqu'au déploiement en production.

En combinant la flexibilité de SageMaker avec d'autres services AWS tels que Lambda Functions et S3, la Banque Efrei peut créer des workflows de machine learning robustes et efficaces pour répondre à ses besoins commerciaux en constante évolution.

3. ETL (Extract, Transform, Load) Informatica :

Les processus ETL sont essentiels pour préparer les données brutes stockées dans le Data Lake pour une analyse ultérieure. Ils extraient les données pertinentes, les transforment selon les besoins métier et les chargent dans des entrepôts de données ou des Data Warehouses.

Voici comment Informatica peut être intégré dans l'écosystème AWS :

- Extraction des données :
 - Informatica peut se connecter à diverses sources de données, y compris les bases de données, les fichiers plats et les API, pour extraire les données brutes stockées dans le Data Lake AWS S3.
 - Il peut également extraire des données à partir d'autres sources externes, telles que des applications tierces ou des systèmes partenaires, pour une intégration complète des données.
- Transformation des données :
 - Une fois les données extraites, Informatica offre un large éventail de fonctionnalités de transformation pour nettoyer, normaliser, enrichir et agréger les données selon les besoins métier de la Banque Efrei.
 - Les transformations peuvent être définies à l'aide d'une interface visuelle conviviale, permettant aux développeurs et aux analystes de données de définir des règles de transformation sans nécessiter de codage manuel.
- Chargement des données :
 - Une fois les données transformées, Informatica peut charger les données dans des entrepôts de données, tels que Snowflake.
 - Il offre des fonctionnalités avancées de gestion de chargement pour garantir l'intégrité et la cohérence des données chargées dans l'entrepôt.

4. Snowflake Data Warehouse :

- Snowflake est un entrepôt de données cloud qui permet de stocker, d'analyser et de partager des données à grande échelle.
- Il offre une architecture multi-cluster, ce qui permet d'évoluer facilement en fonction des besoins de l'entreprise.

- Les données transformées à partir de l'ETL sont chargées dans Snowflake pour une analyse approfondie, la création de rapports et la prise de décisions basées sur des données fiables et cohérentes.

5. **Tableau** pour la Visualisation :

- Tableau est une plateforme de Business Intelligence qui permet de visualiser et d'analyser les données stockées dans Snowflake et d'autres sources de données.
- Les rapports et les tableaux de bord interactifs créés avec Tableau permettent aux utilisateurs de découvrir des insights, de suivre les performances et de prendre des décisions éclairées en temps réel.

Dans ce flux de données, les technologies utilisées sont parfaitement alignées avec les objectifs de la Banque Efrei en matière d'utilisation et de valorisation des données. Elles permettent de collecter, traiter, stocker et visualiser les données de manière efficace, tout en fournissant les insights nécessaires pour soutenir la stratégie commerciale de l'entreprise.

Question 2 :

Voici 5 rôles clés en matière de données indispensables pour devenir data driven :

- **Data Leader** : Développe et met en œuvre une stratégie globale des données, assurant ainsi l'alignement des initiatives de gestion des données avec les objectifs commerciaux de l'entreprise. Il supervise les équipes de données pour garantir la cohérence et l'efficacité dans la gestion des données à l'échelle de l'entreprise. En collaboration avec les parties prenantes internes, il identifie les besoins métier prioritaires et recommande les meilleures pratiques et outils pour la gestion des données.
Il utilise des plateformes de gestion des données telle que Informatica pour la gouvernance et la documentation des données, des outils de collaboration comme Microsoft Teams.
- **Data Architect** : Responsable de la conception globale de l'architecture des données pour garantir que le système de Efrei Banque soit compatible avec la stratégie de données ; par la création de modèles de données, la définition des flux de données, et la garantie que l'architecture permet une intégration efficace et une utilisation optimale des données. Le Data Architect travaillera en collaboration avec les équipes métier pour comprendre les besoins et les traduire en solutions architecturales.
Il utilise les outils de modélisation de données comme ER/Studio, qui permet la conception, la documentation et la gestion des bases de données relationnelles et des entrepôts de données.

- **Data Scientist** : Responsable du développement de modèles prédictifs pour optimiser les offres de produits et services de la banque. C'est lui qui va permettre en partie de réduire le taux de désabonnement de sa clientèle plus jeune, de vendre davantage de prêts immobiliers et d'assurances, puis d'ouvrir de nouveaux bureaux là où cela est pertinent pour la croissance de l'entreprise en fonction des prédictions faites.
Il utilise des outils tels que Python et la plateforme cloud AWS pour l'analyse de données.
- **Data Engineer** : Chargé de la mise en œuvre pratique de l'architecture des données définie par le Data Architect. Ses responsabilités comprennent la collecte, le nettoyage, la transformation et le chargement des données à partir de diverses sources par le biais d'un ETL par exemple. Le Data Engineer sera également impliqué dans la maintenance des pipelines de données et dans la gestion quotidienne des flux de données, veillant à ce qu'ils soient conformes aux normes de qualité et de sécurité définies.
Il utilise des outils comme AWS Cloud S3 bucket, Informatica et Snowflake Data Warehouse.
- **Data Analyst** : Dernier maillon de la chaîne, il est responsable de l'analyse des données finales qui ont été traitées par le Data Engineer et le Data Scientist. Il fournit des dashboards exploitables aux différentes équipes métier afin de les aider à faire et orienter les choix pour Efrei Banque afin qu'ils soient au plus près des attentes de la direction.
Il utilise des outils comme SQL et Tableau pour créer des rapports et des visualisations.
- **Data Governance Manager** : Chargé de mettre en place des politiques et des processus pour garantir la qualité, la confidentialité et la conformité des données définies par le CDO.
Il utilise des outils de gouvernance des données tels que Informatica.

Question 3 :

Une organisation data driven est une organisation axée sur les données . Cela signifie que les décisions stratégiques sont fondées sur l'analyse et l'interprétation des données.

Les raisons stratégiques motivant l'Efrei Bank à devenir Datadriven sont :

- **Amélioration de l'expérience client** : En utilisant les données, issues des formulaires et interactions clients, Efrei Bank pourra mieux comprendre les besoins et comportements de sa clientèle, en particulier des clients plus jeunes. L'analyse des données permettra de segmenter les clients, d'identifier les tendances et de personnaliser les offres et les services pour répondre plus efficacement à leurs attentes..

- **Optimisation des opérations internes** : L'analyse des données peut aider à identifier les inefficacités dans les processus internes et à les améliorer pour une meilleure productivité et des coûts réduits. Cela peut inclure l'automatisation de tâches répétitives, l'optimisation des flux de travail et l'identification des domaines où des améliorations sont nécessaires.
- **Prise de décision basée sur les données** : En s'appuyant sur des données fiables et des analyses approfondies, la direction peut prendre des décisions plus éclairées et stratégiques, ce qui conduit à une meilleure performance globale de l'entreprise. Cela inclut l'allocation de ressources, l'identification des priorités commerciales, et la révision des stratégies de marketing et de vente pour mieux répondre aux besoins du marché.
- **Anticipation des tendances du marché** : En surveillant et en analysant les données du marché, la banque peut anticiper les tendances émergentes et ajuster sa stratégie commerciale en conséquence pour rester compétitive et proposer de nouveaux produits et services.
- **Anticipation et gestion des risques pour assurer la conformité réglementaire** : L'analyse des données peut aider à identifier les schémas de fraude potentiels, à détecter les comportements suspects et à renforcer les mesures de sécurité. En adoptant cette approche, la banque peut mieux répondre aux exigences réglementaires en matière de protection des données et de lutte contre la fraude.

Les difficultés qui pourraient représenter d'énormes défis pour devenir une banque axée sur les données incluent :

- **Intégration et Gestion des Données** : La banque Efrei est confrontée à des défis liés à la fragmentation des données, avec chaque département utilisant ses propres systèmes et sources de données, et à l'intégration de données provenant de différentes sources, ce qui peut être complexe en raison de formats de données incompatibles et de systèmes disparates. Cette fragmentation inclut également la gestion des données non structurées telles que les documents numérisés, les e-mails et les vocaux qui peuvent être difficiles à gérer et à analyser efficacement. Pour surmonter ces obstacles, il est essentiel de mettre en place une stratégie de gestion des données centralisée, utilisant des technologies telles que les entrepôts de données, les plateformes d'intégration des données et des services cloud afin d'assurer une cohérence et une accessibilité accrues des données à travers l'organisation.
- **Confidentialité et sécurité des données** : La protection des données sensibles et la conformité aux réglementations en matière de confidentialité constituent un défi majeur, en particulier avec l'augmentation des cybermenaces. Cela nécessite des investissements dans des mesures de sécurité robustes, ainsi que dans la formation du personnel sur les meilleures pratiques en matière de protection des données et de respect de la vie privée.
- **Culture d'entreprise et adoption** : Le changement de mentalité et la promotion d'une culture axée sur les données au sein de l'organisation peuvent rencontrer des résistances et nécessiter un effort important en termes de sensibilisation et de formation.

- **Qualité des données** : La qualité des données est essentielle pour des analyses précises et fiables. Efrei Bank devra mettre en place des processus de gouvernance des données pour garantir l'exactitude, la cohérence et l'intégrité des données. Cela implique la mise en place de normes de qualité des données, de procédures de nettoyage et de validation des données, ainsi que de mécanismes de suivi et de correction des erreurs.

Pour relever ces défis, Efrei Bank pourrait déployer les cas d'utilisation de données suivants:

- **La segmentation client pour personnaliser les offres et les services**

En utilisant des techniques d'analyse des données, la banque peut segmenter sa clientèle en différents groupes en fonction de critères tels que l'âge, le comportement d'achat, les préférences, etc. En comprenant les besoins et les préférences spécifiques des clients, la banque peut proposer des offres personnalisées pour les produits d'assurance et les prêts immobiliers, ce qui peut augmenter les ventes et fidéliser la clientèle. Par exemple, en identifiant les segments de clients les plus rentables ou les plus à risque de désabonnement.

- **Intégration de données provenant de multiples canaux de communication :**

Ce cas d'utilisation vise à harmoniser les données provenant de divers canaux de communication avec les clients, tels que les réseaux sociaux, les e-mails, les appels téléphoniques et les interactions en personne. En utilisant des outils d'intégration de données robustes, comme Informatica ou AWS Glue, la Banque Efrei peut automatiser le processus d'intégration et de consolidation de ces données provenant de sources disparates.

- **La détection de fraude pour améliorer la sécurité et réduire les risques**

En utilisant des techniques d'analyse avancées, telles que l'apprentissage automatique et l'intelligence artificielle, les banques peuvent identifier les comportements suspects et les transactions frauduleuses en temps réel. Par exemple, en analysant les habitudes de dépenses des clients, les modèles de transactions et les données de localisation, une banque peut détecter des transactions anormales et déclencher des alertes pour une enquête plus approfondie.

- **L'analyse prédictive pour anticiper les besoins des clients et optimiser les processus métiers**

En utilisant des techniques d'analyse prédictive telles que les modèles de machine learning, Efrei Bank peut anticiper les besoins des clients en se basant sur les données historiques et en identifiant les tendances. Par exemple, en analysant les habitudes de dépenses passées d'un client, la banque peut prédire quels produits ou services pourraient intéresser ce client à l'avenir et lui proposer des offres personnalisées en conséquence.

- **La visualisation de données pour améliorer la prise de décision**

Les tableaux de bord de performance fournissent une vue d'ensemble des principaux indicateurs de performance (KPIs) de l'entreprise en temps réel. En utilisant des données en temps réel et des outils de visualisation, la banque peut surveiller et analyser ses performances opérationnelles, financières et commerciales, ce qui lui permet de prendre des décisions rapides et éclairées pour optimiser ses résultats.

Question 4 :

Les trois thèmes qui sont cruciaux pour commencer et établir des bases solides en matière de gestion des données à la Banque Efrei, en garantissant la cohérence, la sécurité et la qualité des données tout au long de leur cycle de vie sont :

Data Modelling & Design (Modélisation et Conception des Données) :

- Domaines de données :
 - Données clients,
 - Données financières.
- Activités :
 - Concevoir et maintenir des modèles de données pour représenter les structures et les relations des données.
Exemple : Concevoir un modèle de données relationnel pour les clients de la banque, décrivant les informations telles que le nom, l'adresse, le numéro de compte, etc.
 - Développer des schémas de base de données optimisés pour répondre aux besoins métier.
Exemple : Créer un schéma de base de données efficace pour stocker les transactions financières, en optimisant l'indexation pour des recherches rapides.
 - Collaborer avec les parties prenantes pour s'assurer que les modèles de données répondent aux exigences fonctionnelles et opérationnelles.
Exemple : Travailler avec les équipes des opérations bancaires pour s'assurer que le modèle de données prend en compte les processus de gestion des comptes clients.

Data Security (Sécurité des Données) :

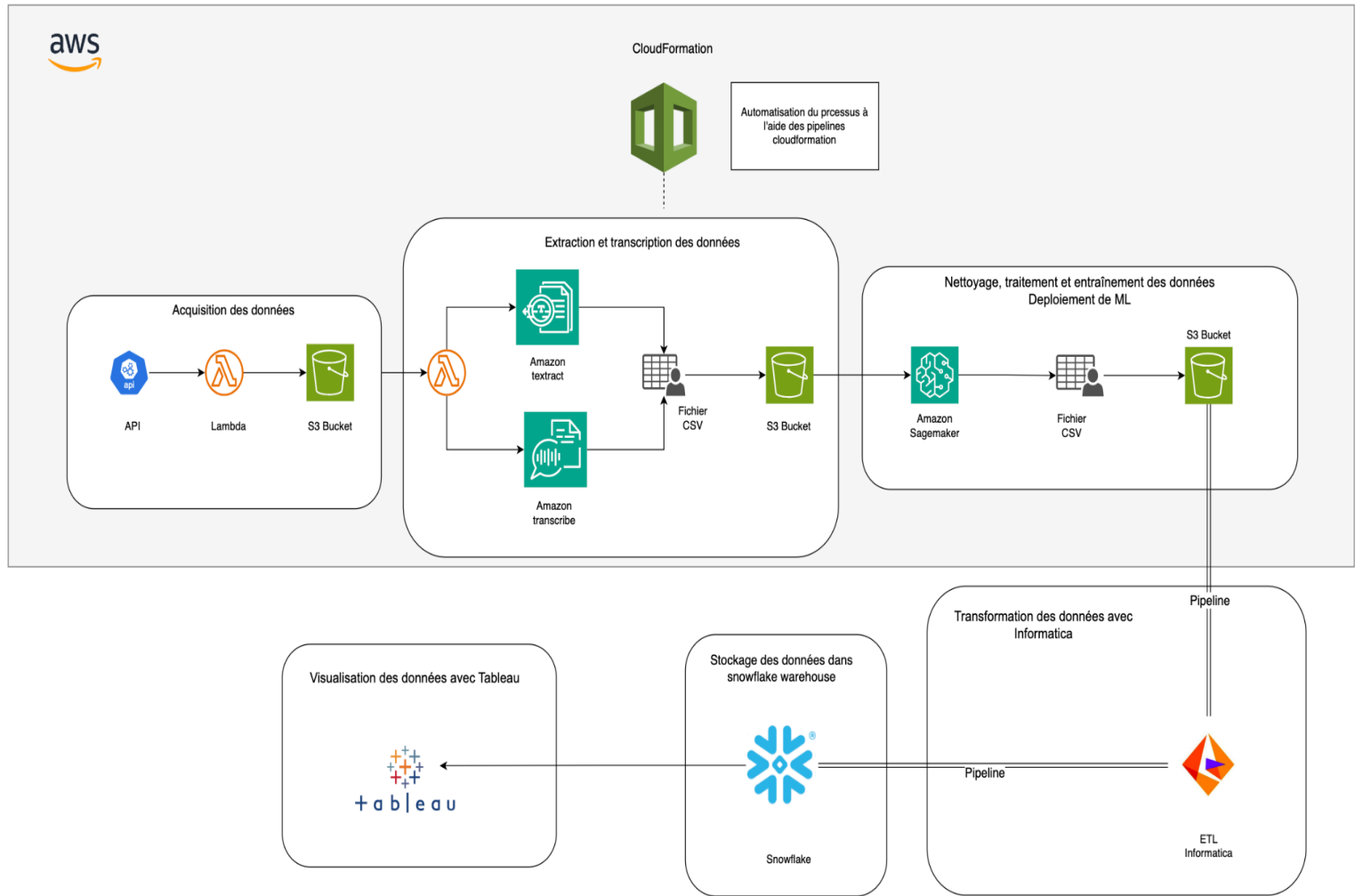
- Domaines de données :
 - Données clients,
 - Données financières.
- Activités :
 - Mettre en œuvre des mesures de sécurité pour protéger les données sensibles contre les menaces internes et externes.
Exemple : Crypter les données clients stockées dans la base de données pour empêcher tout accès non autorisé.
 - Gérer les autorisations d'accès aux données pour garantir que seules les personnes autorisées peuvent y accéder.
Exemple : Définir des rôles d'accès spécifiques dans la base de données, tels que les administrateurs, les utilisateurs réguliers et les auditeurs, avec des niveaux d'autorisation appropriés.
 - Surveiller et auditer l'accès aux données pour détecter les violations de sécurité potentielles et y remédier.
Exemple : Mettre en place des journaux d'audit pour enregistrer les accès aux données et les modifications apportées, et surveiller ces journaux régulièrement pour détecter toute activité suspecte.

Data Quality (Qualité des Données) :

- Domaines de données :
 - Données clients,
 - Données financières,
 - Données marketing.
- Activités :
 - Évaluer et améliorer la qualité des données en identifiant et en corrigeant les erreurs, les doublons et les incohérences.
Exemple : Utiliser des scripts de validation de données pour identifier les entrées incorrectes ou manquantes dans les champs des formulaires clients.
 - Définir et appliquer des normes de qualité des données pour garantir la précision, la cohérence et l'intégrité des données.
Exemple : Établir des règles de validation des données pour s'assurer que les numéros de compte bancaire suivent un format spécifique et respectent les contraintes de longueur.
 - Surveiller les métriques de qualité des données et mettre en place des processus de gestion des anomalies pour maintenir la qualité à long terme.

Exemple : Suivre les tendances de qualité des données à l'aide de tableaux de bord et d'indicateurs clés de performance (KPI), et mettre en place des processus d'assurance qualité pour traiter les problèmes identifiés.

Question 5 :



Voici une vue d'architecture pour le cas d'utilisation "Segmentation clientèle et personnalisation des offres" avec les technologies spécifiées:

1. Acquisition des données :

- ❖ Les données sont collectées à partir de diverses sources telles que les formulaires sur le site web de la banque, les campagnes marketing et les réclamations par téléphone.

- ❖ Les données sont stockées dans un bucket S3 d'Amazon Web Services (AWS) sous forme de fichiers CSV, PDF ou MP3.
2. Extraction et transcription des données:
- ❖ Amazon Textract et Amazon Transcribe sont utilisés pour extraire et transcrire les données à partir des fichiers PDF et des enregistrements audio.
 - ❖ Des fonctions AWS Lambda déclenchent l'API d'Amazon Textract et d'Amazon Transcribe pour extraire les données et les stocker dans des fichiers CSV sur S3.
3. Nettoyage, traitement et entraînement des données :
- ❖ Les données extraites sont nettoyées et traitées à l'aide d'Amazon SageMaker, une plateforme de machine learning d'AWS.
 - ❖ Des modèles de machine learning sont entraînés pour segmenter la clientèle et personnaliser les offres en fonction des besoins et des préférences.
4. Déploiement des modèles de prédiction:
- ❖ Les modèles de prédiction sont déployés sur Amazon SageMaker pour effectuer des prédictions en temps réel sur de nouveaux jeux de données.
 - ❖ Les prédictions sont stockées sur S3 sous forme de fichiers Excel ou CSV pour une utilisation ultérieure dans le processus d'ETL.
5. Transformation des données avec Informatica:
- ❖ Informatica est utilisé comme outil ETL pour transformer les données stockées sur S3 en un format compatible avec Snowflake.
 - ❖ Les données transformées sont chargées dans Snowflake à l'aide de pipelines ETL pour une analyse et une visualisation ultérieures.
6. Visualisation des données avec Tableau :
- ❖ Les données stockées dans Snowflake sont accessibles depuis Tableau pour créer des tableaux de bord interactifs et des rapports détaillés.
 - ❖ Les utilisateurs peuvent explorer les données segmentées et personnalisées, identifier les tendances et prendre des décisions stratégiques basées sur les insights obtenus.

Cette architecture permet à la Banque Efrei de segmenter sa clientèle et de personnaliser les offres en utilisant des technologies de pointe telles que Amazon Textract, Amazon Transcribe, Amazon SageMaker, Informatica, Snowflake et Tableau. Elle assure un flux de données cohérent depuis l'acquisition jusqu'à la visualisation, permettant ainsi une prise de décision stratégique efficace basée sur les données.

Dans le cadre de la segmentation clientèle et de la personnalisation des offres, Efrei Bank peut utiliser l'analyse prédictive pour anticiper les besoins des clients et proposer des offres personnalisées. Cette approche implique d'utiliser différents indicateurs, tels que le nombre de plaintes, le temps moyen passé sur le site web ou l'application, le nombre de visites en agence, et le nombre de produits souscrits, pour prédire les clients les plus susceptibles de se désabonner ou de résilier leur contrat. Ces données peuvent être visualisées dans un tableau de bord avec des graphiques et des indicateurs clés, tels que le taux de désabonnement prévu et le nombre de clients à risque. Cette analyse permet aux équipes

marketing et commerciales de prendre des mesures proactives pour retenir ces clients, telles que des offres personnalisées ou des campagnes de fidélisation ciblées.

Question 6 :

Data Modelling & Design (Modélisation et Conception des Données) :

KPI : Nombre de changements apportés aux modèles de données par rapport au nombre total de modèles.

Exemple : Suivre le ratio entre le nombre de modifications effectuées aux modèles de données, comme l'ajout de nouvelles entités ou l'altération des relations, et le nombre total de modèles existants. Cela permettra d'évaluer la dynamique de l'évolution des besoins métier et la réactivité de l'équipe de modélisation aux changements.

Objectif : Créer des modèles de données pour tous les principaux produits financiers offerts par la banque, y compris les prêts, les comptes d'épargne et les assurances.

Data Storage & Operations (Stockage et Opérations sur les Données) :

KPI : Taux d'utilisation de l'espace de stockage dans le Data Warehouse.

Exemple : Mesurer la proportion d'espace de stockage utilisé par rapport à la capacité totale disponible dans le Data Warehouse.

Objectif : Maintenir le taux d'utilisation de l'espace de stockage en dessous de 80% pour garantir la disponibilité continue de l'espace de stockage nécessaire aux opérations quotidiennes.

Data Security (Sécurité des Données) :

KPI : Nombre de violations de sécurité des données signalées.

Exemple : Suivre le nombre de cas où des accès non autorisés ou des tentatives de piratage ont été détectés dans les systèmes de la banque.

Objectif : Réduire le nombre de violations de sécurité des données de 20% au cours de l'année en cours en renforçant les contrôles d'accès et en sensibilisant le personnel à la sécurité des données.

Data Integration & Interoperability (Intégration et Interopérabilité des Données) :

KPI : Temps moyen d'intégration des nouvelles sources de données.

Exemple : Mesurer le délai entre l'identification d'une nouvelle source de données et son intégration réussie dans le système de gestion des données de la banque.

Objectif : Atteindre un taux de réussite des intégrations de données de 95% d'ici la fin de l'année, en améliorant les processus d'intégration et en investissant dans des technologies d'interopérabilité plus efficaces.

Document & Content Management (Gestion des Documents et du Contenu) :

KPI : Taux d'adoption du système de gestion de contenu par les utilisateurs.

Exemple : Suivre le pourcentage d'employés de la banque qui utilisent activement le système de gestion de contenu pour stocker et partager des documents.

Objectif : Réduire le temps moyen pour retrouver un document de 30% d'ici la fin du trimestre en optimisant les processus de gestion des documents et en mettant en œuvre une classification plus efficace.

Reference & Master Data (Données de Référence et Données Maîtres) :

KPI : Précision des données de référence.

Exemple : Mesurer le pourcentage de données de référence (par exemple, les codes de produit) qui sont correctes et à jour par rapport aux normes définies.

Objectif : Atteindre un taux de complétude des données de référence de 98% d'ici la fin de l'année, en mettant en place des processus de validation des données et en sensibilisant les utilisateurs à l'importance de remplir toutes les informations requises.

Data Warehousing & BI (Entrepôts de Données et Business Intelligence) :

KPI : Temps moyen de génération de rapports BI.

Exemple : Suivre le temps nécessaire pour générer des rapports d'analyse commerciale à partir du Data Warehouse, de la préparation des données à la visualisation des résultats.

Objectif : Réduire le temps moyen de génération de rapports de 25% d'ici la fin du semestre en optimisant les performances de l'entrepôt de données et en automatisant les processus de génération de rapports.

Metadata Management (Gestion des Métadonnées) :

KPI : Nombre de métadonnées documentées par rapport au nombre total.

Exemple : Suivre le pourcentage de métadonnées associées aux ensembles de données et aux modèles de données qui sont documentées de manière exhaustive.

Objectif : Documenter toutes les métadonnées associées à chaque ensemble de données dans le catalogue de métadonnées de la banque d'ici la fin du trimestre pour assurer une documentation complète et précise.

Data Quality (Qualité des Données) :

KPI : Taux de conformité aux normes de qualité des données.

Exemple : Évaluer le pourcentage de données dans le système qui respectent les critères de qualité définis, tels que l'exactitude, la cohérence et l'intégrité.

Objectif : Améliorer le taux d'exactitude des données de 90% à 95% d'ici la fin de l'année en identifiant et en corrigeant les erreurs de données et en améliorant les processus de saisie des données.

Data Architecture (Architecture des Données) :

KPI : Nombre de projets conformes à l'architecture des données définie.

Exemple : Suivre le nombre de projets de développement ou de migration qui respectent les principes et les normes architecturaux établis pour garantir la cohérence et la compatibilité du système dans son ensemble.

Objectif : Intégrer avec succès cinq nouveaux systèmes ou applications au réseau de données de la banque d'ici la fin de l'année pour renforcer l'architecture des données et répondre aux besoins croissants de l'entreprise.