**Conception SID**

**Cahier des Charges**

**-**

**Document fonctionnel**

**Table des matières**

[1. Introduction 3](#_Toc186813555)

[2. Besoins utilisateurs 4](#_Toc186813556)

[2.1 Données des vols des participants 4](#_Toc186813557)

[2.2 Comportements des participants 4](#_Toc186813558)

[2.3 Promotion sur les vols 4](#_Toc186813559)

[2.4 Canaux de distribution 4](#_Toc186813560)

[2.5 Monitoring et mesures clés 4](#_Toc186813561)

[3. Système ETL 5](#_Toc186813562)

[3.1 Fonctionnalités ETL - spécifications 5](#_Toc186813563)

[3.2 Qualité des données 6](#_Toc186813564)

[4. Datawarehouse 8](#_Toc186813565)

[5. Gestion projet 9](#_Toc186813566)

[5.1 Méthodologie 9](#_Toc186813567)

[5.2 Livrables attendus 9](#_Toc186813568)

# Introduction

Le projet concerne un programme marketing pour les passagers réguliers d'une compagnie aérienne. L’objectif est d’analyser les vols de chaque participant pour mieux comprendre leur comportement et optimiser les offres, tout en assurant un suivi des miles, des surclassements et d'autres indicateurs de performance. Ce projet s'inscrit dans une stratégie de fidélisation et d'optimisation des revenus de la compagnie aérienne.

Pour ce faire, on souhaite mettre en place une infrastructure de « Business Intelligence ».

Cette infrastructure se composera des éléments suivants :

### DataWarehouse

Ce DataWarehouse prendra la forme d’une base de données relationnelle.

Son rôle sera de consolider en un point central les données fonctionnelles de

l’entreprise, de manière homogène (formats identiques pour des données de même type), exhaustive (consolidation de toutes les applications fonctionnelles) et qualitative (vérification de la qualité des données stockées).

### Système « ETL » (Extract-Transform-Load)

En amont du DataWarehouse, le système ETL permettra de :

* + Extraire (Extract) les données des différentes sources de données fonctionnelles.
  + Transformer (Transform) ces données pour atteindre les règles d’homogénéité et de qualité voulues.
  + Charger (Load) ces données dans le DataWarehouse

# Besoins utilisateurs

Les utilisateurs de l'application souhaitent disposer d'une solution permettant une analyse approfondie et un suivi précis des comportements des participants au programme marketing. Cette analyse doit couvrir une période de 3 ans d'historisation et inclure les dimensions suivantes :

## Données des vols des participants

* Identification des vols pris : Liste et suivi des vols empruntés par les participants, incluant les trajets, les dates, et les compagnies aériennes.
* Analyse des appareils utilisés : Étude des types d'appareils (modèles d'avions) sur lesquels les participants voyagent, afin d'identifier les préférences et tendances.
* Étude des tarifs de base : Analyse des tarifs payés par les participants pour chaque vol, permettant de mieux comprendre les habitudes de dépenses.

## Comportements des participants

**Fréquences des surclassements** : Suivi des habitudes de surclassement des participants.

**Miles offerts et consommés** : Monitoring des miles attribués (offres, promotions) et de leur utilisation par les participants.

## Promotion sur les vols

**Analyse de l'efficacité des promotions** : Étude des périodes et types de promotions qui génèrent le plus d'engagement et d'achats parmi les participants.

## Canaux de distribution

**Analyse des ventes par canal** : Ventilation des ventes selon les différents canaux (vente directe, agences de voyage, plateformes en ligne, etc.) pour identifier les canaux les plus performants.

## Monitoring et mesures clés

* Taux de conversion des vols : Calcul du taux de conversion pour chaque canal de distribution, défini comme le ratio :
* (Nombre de visiteurs via un canal / Nombre d’achats via ce canal).
* Durée des séjours : Suivi de la durée moyenne des séjours des participants après leurs vols, en lien avec les habitudes de voyage.
* Analyse des classes de voyage : Suivi de la répartition des participants par classes (économique, premium, business, première) pour identifier les tendances.

# Système ETL

## Fonctionnalités ETL - spécifications

### Extract

Le système ETL devra permettre d’extraire des données provenant des sources suivantes :

* + - *Fichiers csv*

Nous avons généré 13 fichiers sources csv pour les données de notre projet.

### Transform

Une fois les données extraites, le système ETL permettra de les transformer via Talend.

Il devra entre autres être possible de :

* + - Faire correspondre les champs des tables sources et destinations (mapping entre les sources et le DataWarehouse).
    - Filtrer des enregistrements.
    - Filtrer des colonnes.
    - Joindre les informations venant de tables différentes vers une même destination.
    - Transformer les types de champ entre entrée et sortie.
    - Appliquer des règles de formatage entre les données d’entrée et de sortie. Entre autres :
      * format standardisé de date,
      * de numéro de téléphone

### Load

Les données extraites et transformées seront ensuite chargées dans le DataWarehouse.

## Qualité des données

Le système ETL devra permettre de contrôler la qualité des données stockées dans le DataWarehouse et de veiller au maintien de cette qualité.

### Contrôle des données sources

Le système devra disposer d’un outil permettant de juger de la qualité des données sources avant leur intégration dans le DataWarehouse.

Cet outil de contrôle de qualité doit permettre de tirer des statistiques sur les fichiers sources, selon des règles fixes ou configurables (ex : détecter la présence dans un fichier Excel d’IDs manquants ou doubles, alors que l’ID est obligatoire et unique une fois intégré dans la base de données DataWarehouse).

### Contrôle et actions en cours de routine

Durant la routine ETL, le système doit être capable de détecter des problèmes de qualité éventuels et d’agir en conséquence, notamment au niveau :

* *Doublons*

L’algorithme de détection des doublons doit être flexible et intelligent :

* + Prise en compte de combinaisons de champs configurables (nom, prénom, numéro de téléphone, …)
  + Prise en compte de correspondances proches (« like ») et pas uniquement de correspondances exactes (ex : dupond<>dupont)

Les doublons détectés doivent pouvoir être gérés suivant des règles configurables.

En cas de conflits (informations différentes pour un même champ), les actions suivantes peuvent être prises :

* + Action automatique : le système concatène (merge) les doublons suivant une règle établie sans intervention humaine
  + Action manuelle : le système transmet la liste des doublons à un utilisateur qui décide quelles informations sont à conserver/écraser
* *Incohérence de types de données*

Exemple : import d’une valeur décimale dans un champ entier

* *Incohérence de format*

Exemple : import d’un mauvais format de date

* *Incohérence de valeurs*

Exemple : import dans un champ String d’une valeur « A » alors que les seules valeurs supportées par l’application sous-jacente sont « B » ou « C ».

### Contrôles et actions a posteriori

Il doit également être possible de « sonder » ponctuellement le DataWarehouse pour vérifier la qualité des données et prendre les actions nécessaires.

Les algorithmes de contrôles et d’actions décrits ci-dessus doivent donc également être disponibles dans le cadre de contrôles ponctuels.

# Datawarehouse

# Gestion projet

## Méthodologie

Le soumissionnaire devra proposer une méthodologie de gestion de projet itérative et collaborative, garantissant un contrôle régulier et une validation des étapes clés tout au long du projet. Cette méthodologie s’appuiera sur le modèle du « cycle en V », tout en intégrant des approches itératives pour maximiser la flexibilité et l'adaptabilité.

Les étapes suivantes doivent être respectées au minimum :

* **Analyse des besoins** : Identification des besoins fonctionnels et non fonctionnels des utilisateurs, en étroite collaboration avec les parties prenantes.
* **Spécifications fonctionnelles et techniques** : Élaboration des spécifications détaillées décrivant les fonctionnalités et les exigences techniques du projet.
* **Développement** : Réalisation des composants techniques, incluant l’intégration et les tests unitaires.

## Livrables attendus

Les livrables devront refléter une documentation claire, exhaustive et conforme aux standards. Les éléments suivants sont attendus :

* **Documentation fonctionnelle** :
  + Un document fonctionnel succinct pour les spécifications initiales (mini cahier des charges, mini spécifications détaillées).
  + Un effort particulier devra être consacré aux spécifications fonctionnelles pour assurer une compréhension partagée des besoins.
* **Dictionnaire de données ou contrat d'interface** :
  + Une documentation précisant les structures de données, avec une préférence pour un contrat d’interface lorsque cela est pertinent.
* **Matrice décisionnelle** : Un tableau structurant les critères de décision pour valider les choix techniques ou fonctionnels.
* **Modèle conceptuel de données (MCD)** : Représentation des données et de leurs relations pour assurer une modélisation claire et cohérente.
* **Livrables techniques spécifiques** :
  + Livrables liés aux technologies et outils utilisés, notamment pour les projets Talend.