**Projet**

**Client :**

La Postale (www.laposte.fr)

La Poste est un grand groupe de services organisé en cinq branches d’activités : La branche Services - Courrier - Colis, La Banque Postale, Geopost (colis-express), Le réseau La Poste (les bureaux de poste), et Le numérique. Le Groupe La Poste réunit plus de 251 000 collaboratrices et collaborateurs animés par des valeurs citoyennes qui sont depuis toujours au cœur de son ADN et de l’identité postale. Avec un réseau de plus de 17 000 points de contact sur tout le territoire, Le Groupe La Poste est le 1er réseau commercial de proximité en France. À l’international, il renforce ses positions, avec une présence dans plus de 40 pays sur cinq continents. Depuis janvier 2011, 100% du chiffre d’affaires du Groupe est réalisé sur des marchés en totale concurrence. En dix ans, Le Groupe La Poste s’est modernisé en profondeur en intégrant les dimensions numérique et écologique sur un marché concurrentiel.

**Date du projet :**

* Durée de la mission : 18 mois : Janvier 2021-JUIN 2022
* Adresse : Siège social du Groupe La Poste Principal 9 rue du colonel Avia, Paris, 75015

**Nature du projet :**

Mise en place d’un cluster Big Data « Butia » => DLG « Data Lake Groupe La Poste »

Et évolution du Reporting (TABLEAU)

**Le contexte :**

Modernisation et de transformation numérique du groupe et développe de nouveaux services pour le compte des autres branches du groupe.

Des usines à trier le courrier, qui gèrent chaque année 12 milliards de lettres et 300 millions de colis, générant 30 milliards de traces, des données issus soit du suivi des objets de bout en bout, soit de leur traitement sur des robots de tri. Les données n’étaient pas valorisées à part au sein d’études ponctuelles

Un usage massivement répandu au sein de la société, puisque 2 400 utilisateurs ont accès aux tableaux de bord construits avec Tableau Software.

**Objectif du projet :**

Utiliser la donnée pour induire un changement des mentalités

Améliorer l’efficacité opérationnelle de tableau Software.

Améliorer le pilotage opérationnel des plates-formes industrielles pour optimiser le tri du courrier

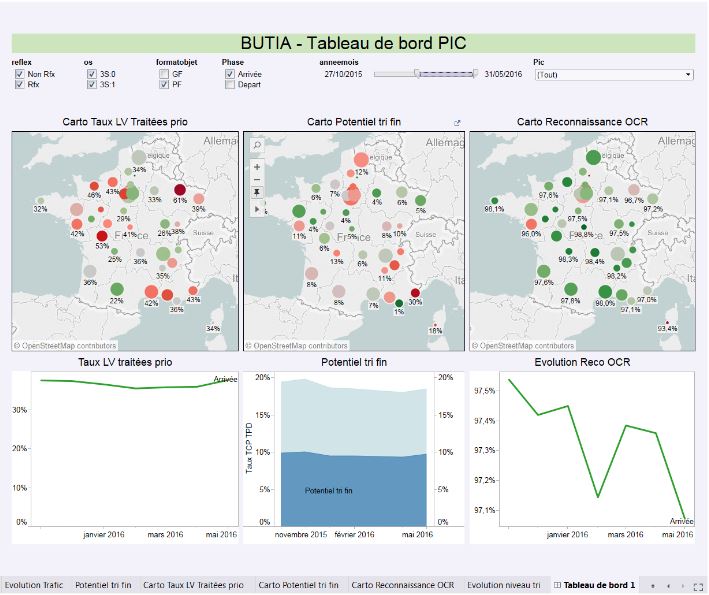
Mise en place d’un Data-Lake pour le Groupe La Poste

**Environnement technique :**

Java8, Tableau, Spark, Kafka, HDFS, Shell, Hive, kibana, Elasticsearch, Sql Postgres

GIT, JENKINS, NEXUS, ANSIBLE

**Définition des KPIs (Key Performance Indicator) :**



**Méthode de travail : Le Poste : Développeur BIG DATA**

**Agile** :

**Chef de projet (Scrum Master/)** : Laurent. D, directeur de l’innovation et du pilotage opérationnel de la performance au sein de la branche Courrier et colis de La Poste

**Product Owner** AMOA, Expert Métier

**05 Collaborateur** (Chef de projet, 2 Développeurs (moi & un développeur BI)

**DevOps :** Outils : GIT, JENKINS, NEXUS, ANSIBLE

**L’architecture Technique du projet :**

A chaque fois qu’on modifie le statut d’un courrier ou d’un colis (par les facteurs et les distributeurs de la poste), une API Rest s’occupe de modifier le fichier XML dans le web

Grâce au procédé technologique OCR « Optical Character Recognition » qui permet de récupérer le texte sur une image (photo ou document numérisé) et de donner en retour un document textuel, exploitable et modifiable, là où un scanneur ne fera que dématérialiser l’image.

API REST DES CLIENTS (destinataires et expéditeurs) DU GROUPE LA POSTE

XML

API REST DES UTILISATEURS (LA POSTE)

Facteurs, employés, etc

KAFKA

SPARK STREAMING

**HDFS HDFS**

Chaque bureau de poste, ou centre de tri pourra visualiser à J+1 l’état d’avancement du tri & distribution de ses flux de courriers et colis

SPARK

Elasticsearch

Kibana

Hive

TABLEAU

Elastic search utilise les données sous forme d(indice)

Système décisionnel

Système opérationnel

**LES TACHES :**

* Concevoir et s’assurer du fonctionnement de l’architecture Big Data de la structure
* Valorisation des données par enrichissement
* Crée & intégrer un job Spark qui va s’occuper de la lecture et la reconnaissance des caractères à partir d’un fichier OCR,
* Assurer une veille technologique.
* Contribuer au développement et à la maintenance du cœur technique sur lequel s'appuient les solutions applicatives de l’entreprise.
* Aide à la construction des différentes couches de données dans HDFS (RAW DATA, INGESTION, TRAITEMENT, EXPOSITION DE RESULTATS)