



**data
iku**



***Comment mettre en production et suivre la performance
de nombreux modèles dans le temps ?***

Retour d'expérience AVISIA * Bouygues Telecom

DATA, DIGITAL & TECHNOLOGY

Fondé en 2007, AVISIA est UN ACTEUR DE RÉFÉRENCE dans le DOMAINE DE LA DATA.

A travers une équipe de PASSIONNÉS et d'experts, nous sommes spécialisés dans le CONSEIL, l'INTÉGRATION et la RÉALISATION de projets DATA CENTRIC.



Retour d'expérience réalisé au sein du pôle connaissance clients intervenant sur des sujets compréhension du comportement client par la valorisation de la donnée de manière transverse.

Partenaire depuis 2009, nous avons su être tour à tour présent aussi bien ***dans de l'accompagnement en renfort d'activité*** pour répondre aux besoins fonctionnels que dans ***la proposition de démarches innovantes génératrices de valeur.***

OBJECTIFS



CONTEXTE / ENJEUX

- » Créer un score
- » Le mettre en production
- » Généraliser, itérer, « Industrialiser »



TEAM

8

Data Analyst

~40

Jours-hommes
pour ce projet
d'industrialisation



OUTILS





CONTEXTE / ENJEUX

- » Créer un score
- » Le mettre en production
- » Généraliser, itérer, « Industrialiser »



Design

Visual Machine Learning

Réduction des variables

Algorithmes





CONTEXTE / ENJEUX

- » Créer un score
- » Le mettre en production
- » Généraliser, itérer, « Industrialiser »



Metrics

Checks

Scenarios





CONTEXTE / ENJEUX

- » Créer un score
- » Le mettre en production
- » Généraliser, itérer, « Industrialiser »

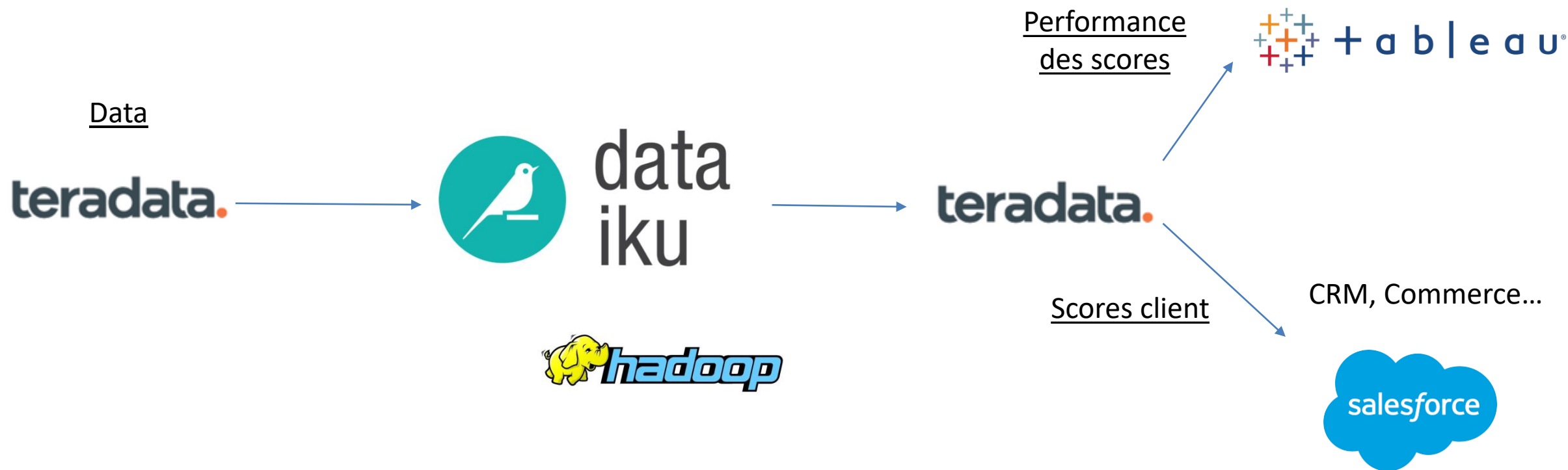


Variables projets

Plugin

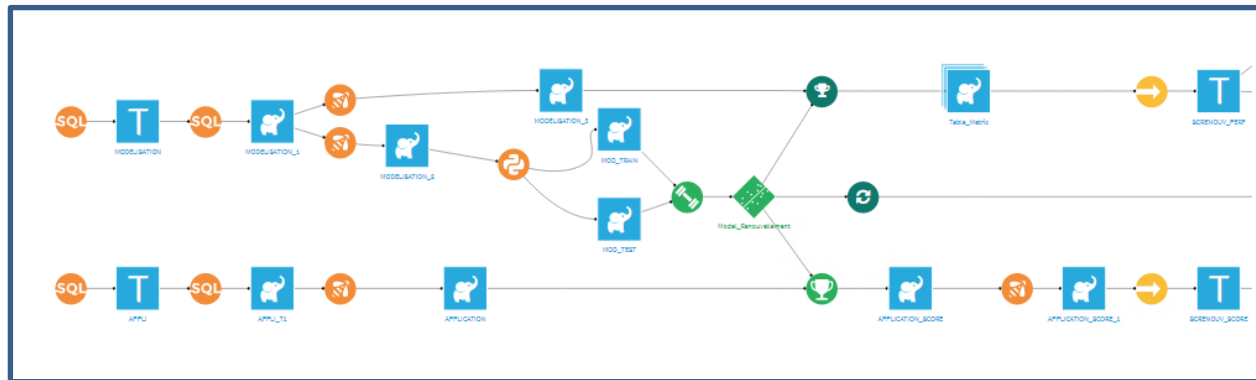
Workflow répliquable pour tout prochain score







Workflow générique



Préparation
des données
dans
Teradata

Reporting : Performance des scores



Scores client



Performance du score : modèle « en production » versus modèles concurrents !

% des clients ciblés

10

Légende

■ Modèle en prod

■ 20200302

■ 20200201

■ 20200101

■ 20191213

■ 20191001

■ 20190901

■ 20190801

■ 20190701

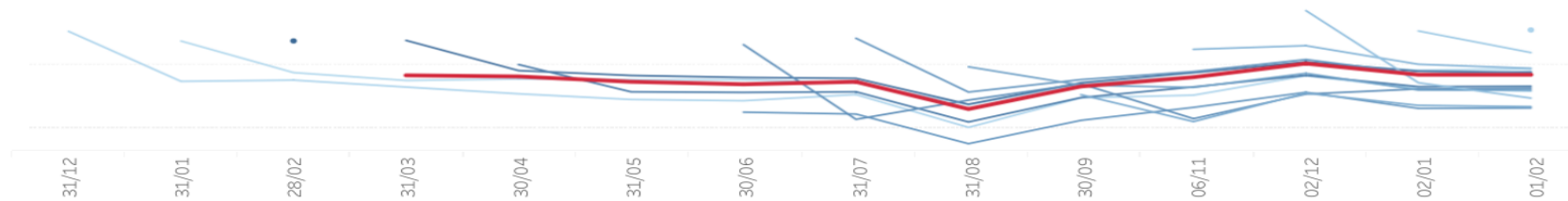
■ 20190601

Date modèle en prod

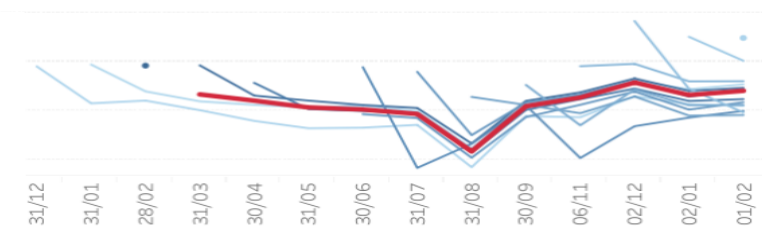
Nombre de modèles
utilisés lors du dernier
calcul

14

AUC

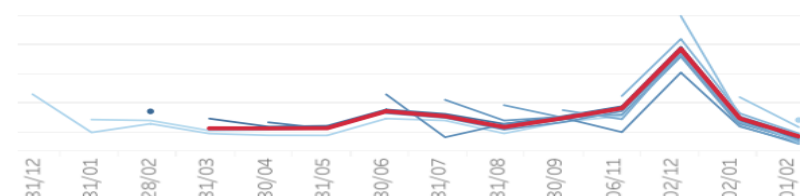


Lift



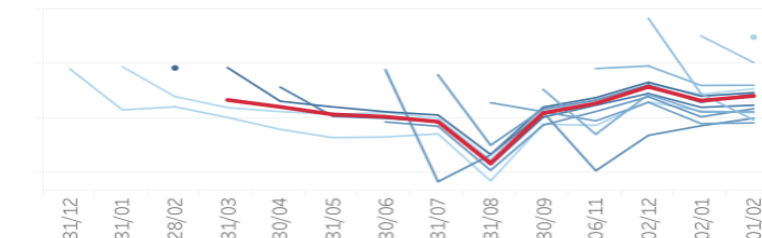
Précision

% des ciblés réellement churners (resp. souscripteurs)



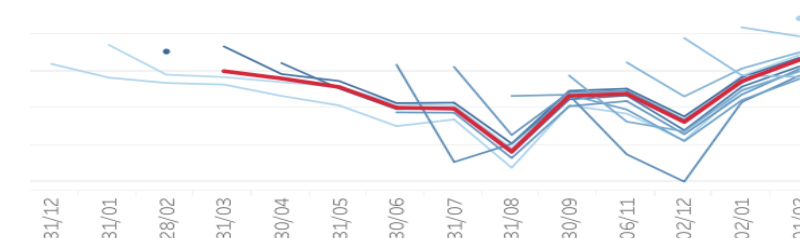
Sensibilité

% de churners (resp. souscripteurs) bien scorés

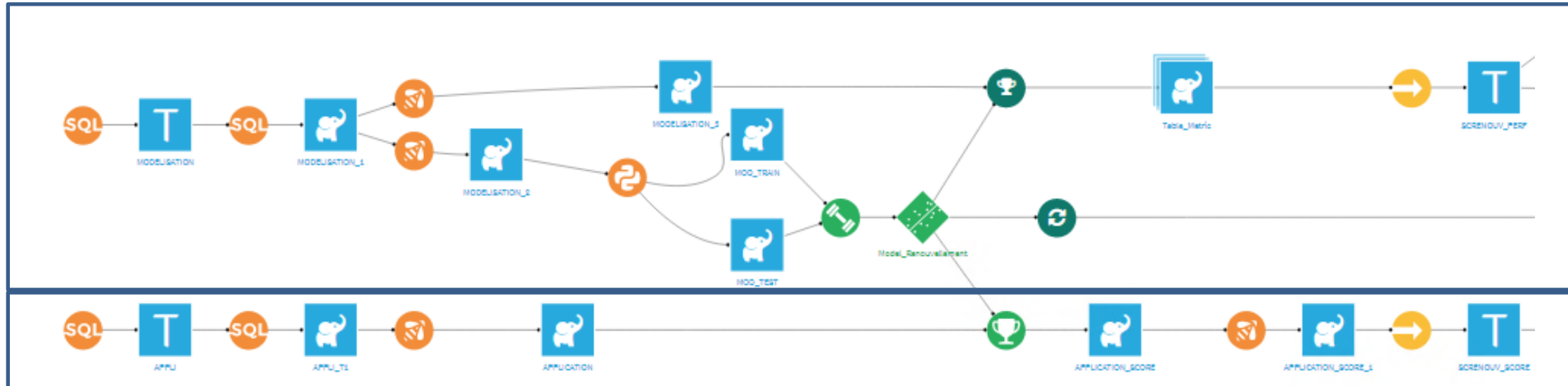


Bien classés

% de clients bien scorés



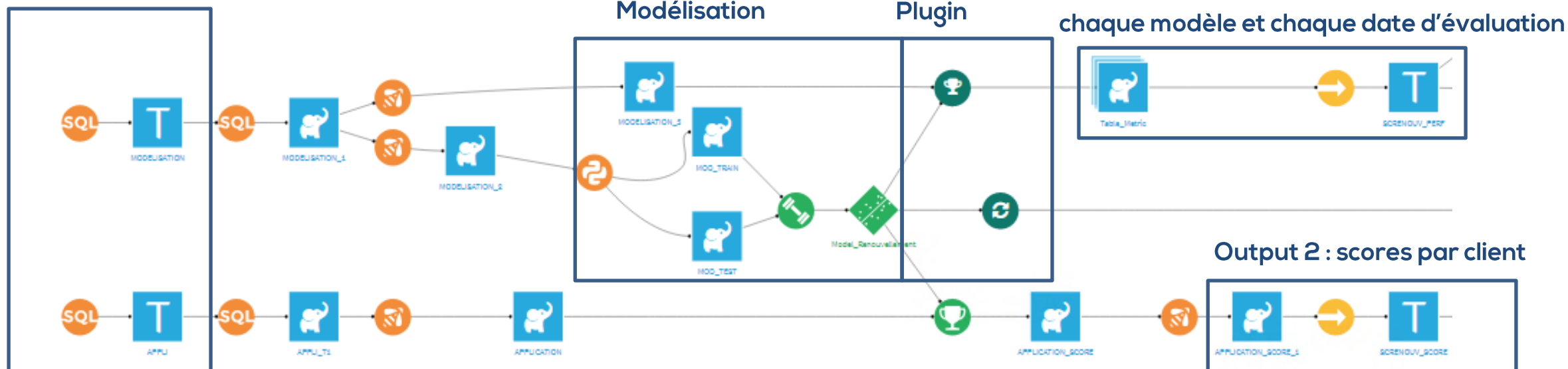
Modélisation : Entraînement, réentraînement, évaluation des modèles entraînés par le passé



Application du score

INPUTS / OUTPUTS & SCENARIOS

Dataprep :
lancement de
procédures Tera



2 scénarios :

☐ Application Renouvellement
(id: Application_Renouvellement)

★ Auto-triggers ☒ ON
👁 Appli_mercredi_jeudi

☐ Score Renouvellement
(id: Score_Renouvellement)

★ Auto-triggers ☒ ON
👁 1er_jour_mois

ORDONNANCEMENT



- ☐ Set project variables date_ref_appli ...
- ☐ Compute metrics Compute Cockpit ...
- ☐ Custom Python Var_cockpit ...
- ☐ Send message Mail COCKPIT erreur P ...
- ☐ Build Procédure Tera P ...
- ☐ Compute metrics Compute procédure P ...
- ☐ Custom Python Var_procedure P ...
- ☐ Send message Mail procédure erreur P ...

Variables de scénario (par ex : heure de début de calcul, heure de fin de calcul)

Alerte mail (échec, temps de calcul)

Alerte mail (échec, temps de calcul)



Exécution chaque semaine

Triggers

ADD TRIGGER

Time-based Appli_mercredi Active ON

Frequency On certain days of the week

Run on days of week Wednesday Days of each week to run the scenario on

Run at 11:40 AM Hour of day at which the scenario will run on the selected day

ORDONNANCEMENT



☐ Set project variables date_ref_appli ...

☐ Compute metrics Compute Cockpit ...

☐ Custom Python Var_cockpit ...

☐ Send message Mail COCKPIT erreur P ...

☐ Build Procédure Tera P ...

☐ Compute metrics Compute procédure P ...

☐ Custom Python Var_procedure P ...

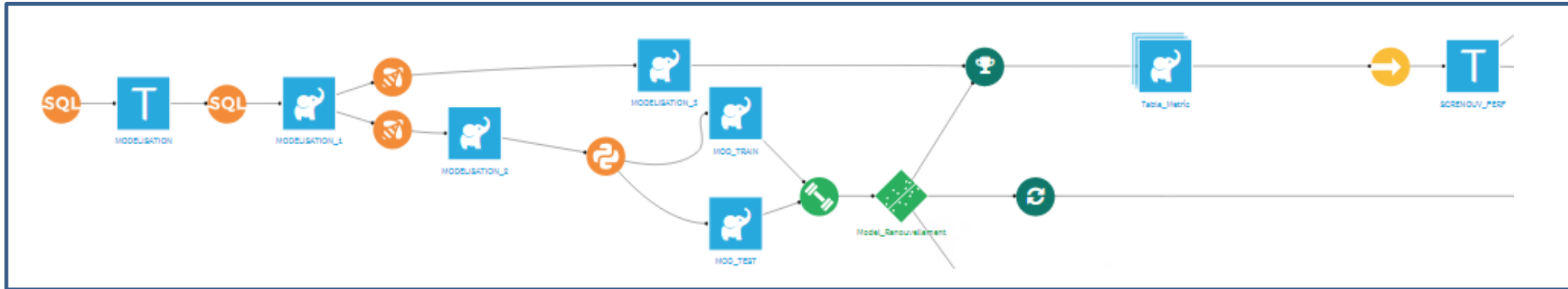
☐ Send message Mail procédure erreur P ...

Custom Python step pour récupérer des metrics ou checks issus d'un autre projet !

Exécution conditionnelle des steps suivants (alerte mail ou exécution des steps suivants)

Custom Python step pour récupérer des metrics dans 2 datasets distincts et créer un check
« manuel » (stabilité des volumétries)

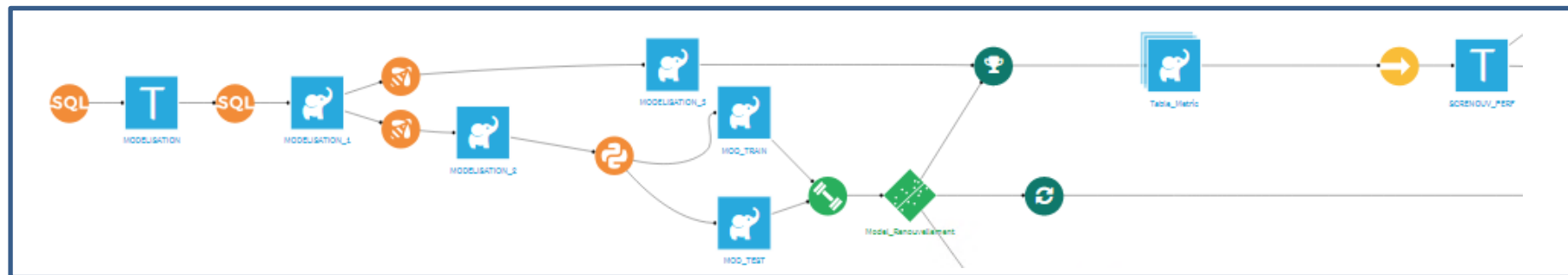
REENTRAINEMENT & MONITORING



Scénario 2 : réentraînement + évaluation des modèles entraînés par le passé

- » Entraînement du même algo, sur de nouvelles données (pas d'activation !)
- » Evaluation (prédiction vs réalité) des performances de tous les modèles !
- » Stockage de toutes les performances (dataset partitionné), alimentation d'un reporting Tableau
- » L'activation d'un modèle challenger reste à la main de l'utilisateur !

REENTRAINEMENT & MONITORING : PLUGIN

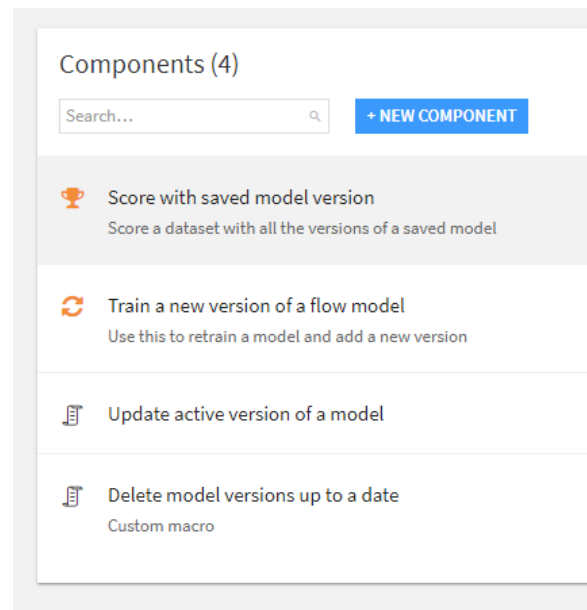


Pourquoi un plugin ?

- » Entraînement du même algo / mêmes hyperparamètres **mais sur de nouvelles données avec un schéma qui peut avoir changé !!** (GUESS)
- » Surtout : des metrics de performance étendues (lift 1, 5, 10%...) et **une évaluation sur une liste évolutive de « modèles à comparer »**

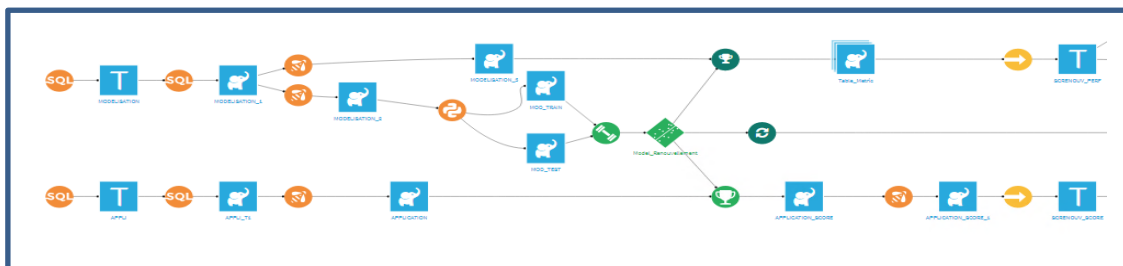
Comment ?

- » Python recipe puis transformation en plugin
- » Ecriture du plugin par un Data Scientist, réutilisation par tous les utilisateurs

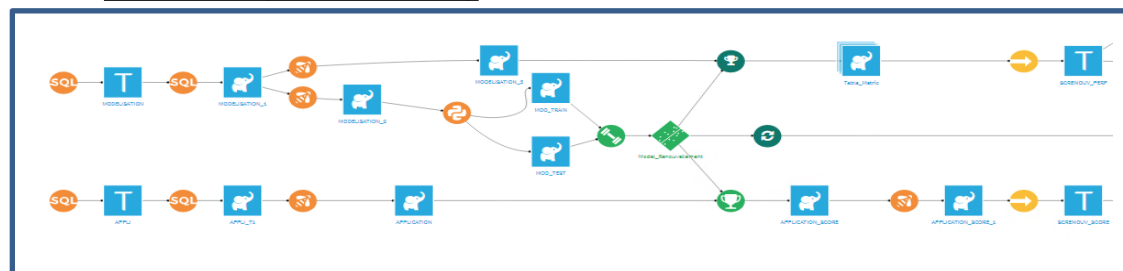




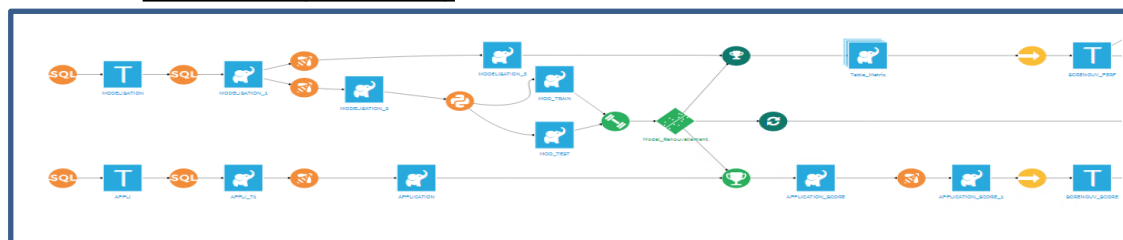
Score 1 (renouvellement)



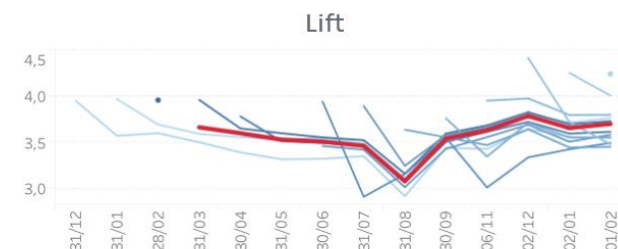
Score 2 (attrition...)



Score 3 (x sell...)

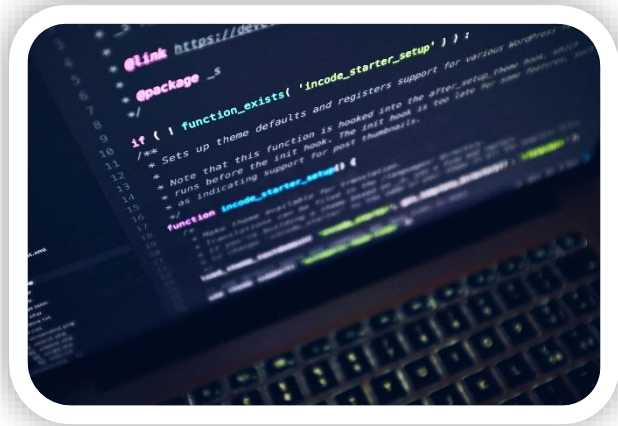


Reporting : Performance des scores



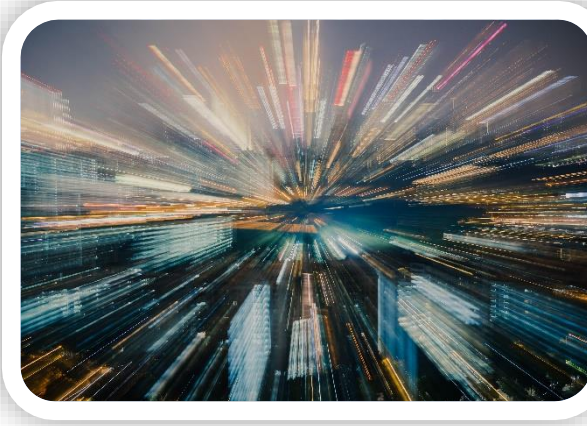
Scores client





Gain de productivité et d'autonomie

- » Directement reproductible pour créer et mettre en production un prochain score
- » Déploiement et suivi à la main d'une équipe de Data analyst



Perspectives de gains en termes de performance de modèle

- » Se réappropriier et challenger les modèles des scores existants
- » Opportunités du Visual ML



Maîtrise du temps de traitement

- » Perspectives pour scaler sur un nombre encore plus important de modèles offertes par Kubernetes

READY FOR TAKEOFF



Pour plus de détails et échanger sur la partie technique & algorithmique, nous serons présents sur le stand A2 & via la plateforme en ligne :

<https://www.connectbycorp.com/big-data-et-ai-paris-2020/>





DATA, DIGITAL & TECHNOLOGY