

## Validation des données

### Description

Après un objet INPUT, les données d'entrée doivent être validées. Les données qui ont été validées dans un script précédent n'ont pas besoin d'être validées à nouveau. La validation détecte les erreurs de données avant qu'elles ne deviennent problématiques et devrait soit produire des avertissements pour des données d'entrée incorrectes ou écrire les données incorrectes dans un fichier de sortie séparé. Un script standard est écrit pour aider à valider.

Vérifiez que les champs numériques ne contiennent pas de lettres. Ceci peut être fait comme suivant :

```
if(<nom_de_champ> + 0 = <nom_de_champ>, "ok", "erreur")
```

Vérifiez que les champs de valeurs date contiennent actuellement une date :

```
if(format_date(<nom_du_champ>, "YYYY-MM-DD")>1, "ok", "erreur")
```

Vérifiez que les champs qui sont supposés être uniques (par exemple les identifiants clients) sont effectivement uniques. Ceci peut être effectué dans un objet SQUASH : en ajoutant une colonne de comptage et en vérifiant si elle contient des valeurs supérieures à 1.

Vérifiez la longueur des champs pour avoir une longueur fixe:

```
if(length(<nom_du_champ>=<longueur>, "ok", "erreur")
```

Un champ chaîne de caractères peut contenir des données numériques, il peut être transformé afin de s'assurer qu'il ne sera pas traité comme un nombre. Un simple calcul dans ce but est :

```
ltrim(concat(" ", <nom_du_champ>))
```

Toutes les dates utilisées dans les calculs doivent être converties en valeur de date (datevalue).

La suppression des colonnes non nécessaires provenant du flux de données est considérée comme un élément de ce processus de validation.

### Tags

1. Data Integrator
2. script