

GROUPE ASKIDA

# RAPPORT

## Analyse BD SQL Hydro Soft

---



### Introduction

Le Groupe Askida® fut choisi pour améliorer un produit de la compagnie Hydro Solutions. Ce document est un effort fait afin de mieux comprendre l'application existante. Son squelette est une BD Sql Server 2012© et utilise Great Plains©. Des coquilles existent, dans les grandes lignes, c'est due au manque de compartimentalisation. Tout réside dans le même schéma de Great Plains©, avec dédoublement et copie de tables, en plus des tables propres à l'application. Les agents externes causent beaucoup de transactions donc trop de pression sur la BD.

---

---

## **Serveur SQL 2012**

L'analyse ne reflète pas la réalité, car nos tests se font dans l'environnement de développement fait par HydroSoft. Il faudra répéter l'analyse de DB dans l'environnement de production pour avoir la surcharge réelle du programme.

### **Schéma de BD**

Les nombreuses transactions entre les tables originales de Great Plains© et les tables HydroSoft (de l'application et en lien avec Great Plains©) causent la surutilisation de mémoire et d'espace disque. Il faudra découpler les nouveaux développements dans un autre Schéma de données (i.e. HydroSoft).

## **Base de Données (BD)**

### **Espace de disque et mémoire**

L'environnement de développement auquel nous avons accès, est assez léger car il consomme presque 900 Mo de mémoire mais prends 620 Go sur disque en plus de résider dans un fichier physique MDF, ceci n'est pas optimal.

### **Contraintes**

Décompte du nombre de contraintes par table, les tables ayant plus de 7 contraintes par table appartiennent à Great Plains©, donc les tables de HydroSoft moins lourdes.

### **Utilisation d'objets**

L'environnement de développement nous montre aussi que la BD est centrée sur des procédures stockées plus que sur des fonctions.

### **Nombre d'enregistrements pour chaque table**

Il y a au-delà de 1000 tables qui n'ont pas d'enregistrements (non utilisées). On estime à 150 tables le nombre de tables de configuration ayant entre 1 et 10 enregistrements. Les

---

table ayant plus de 1000 enregistrements sont les plus utilisées tandis que celles ayant de 100 à 1000 sont les plus liées à la logique d'affaires.

## **Les fonctions**

Plus de 300 fonctions sont présentes. La minorité sont "valuées" retournant un ensemble de données suite à une sélection. La plupart sont "scalaires" retournant une seule valeur. Les fonctions sont optimisées pour diminuer la quantité du flux de données.

### **Les paramètres des fonctions**

Toutes les fonctions reçoivent des valeurs en paramètre.

### **Distribution du nombre de paramètres des fonctions**

Dans la minorité des "valuées", la plupart ont 4 paramètres mais moins que 10, ce qui vans dans le sens de la norme. Par contre dans la majorité des "scalaires", la plupart ont 2, il faudrait investiguer les fonctions ayant de 10 à 20 paramètres pour des possibles optimisations.

## **Les procédures stockées**

Le système comporte au delà de 17 000 procédures stockées, dont la plupart n'utilisent pas de paramètres.

### **Les paramètres des procédures stockées**

Presque 8 000 procédures ont des paramètres.

### **Distribution du nombre de paramètres des procédures stockées**

Il faudrait vérifier celles ayant plus de 50 paramètres ainsi que le groupe de procédures comprises dans le pic de 40 paramètres.

---

## **Les vues**

Plus de 250 vues sont présentes. La moitié dédiée à Great Plains et l'autre moitié pour HydroSoft. La plupart sont "scalaires" retournant une seule valeur. Il faudrait profiler la charge de données sur les vues dans l'environnement de production pour connaître l'empreinte en mémoire et le ratio de lecture/ écriture sur disque.

## **Les indexes**

Les tables ayant moins de 5 index par table doivent être vérifiées.

## **Les clefs principales**

Il resterait à valider 20% des tables ayant moins de 2 clefs par table.

## **Les clefs étrangères**

Il resterait à valider les tables n'ayant pas de 2 clefs vers les principales table.

## **Les tables**

Après avoir fait le survol des autres composantes, des indices clairs existent pour continuer l'exploration des tables et de les modifier. Il faudra refaire l'exercice sur l'environnement en production pour la cueillette de données sur le comportement de la BD.

Il faut développer la nouvelle application dans un nouveau schéma, valider les contraintes, investiguer les tables ayant de 100-1000 enregistrements, valider les accès disque des tables appelées le plus souvent par les procédures, par les fonctions et par les vues.

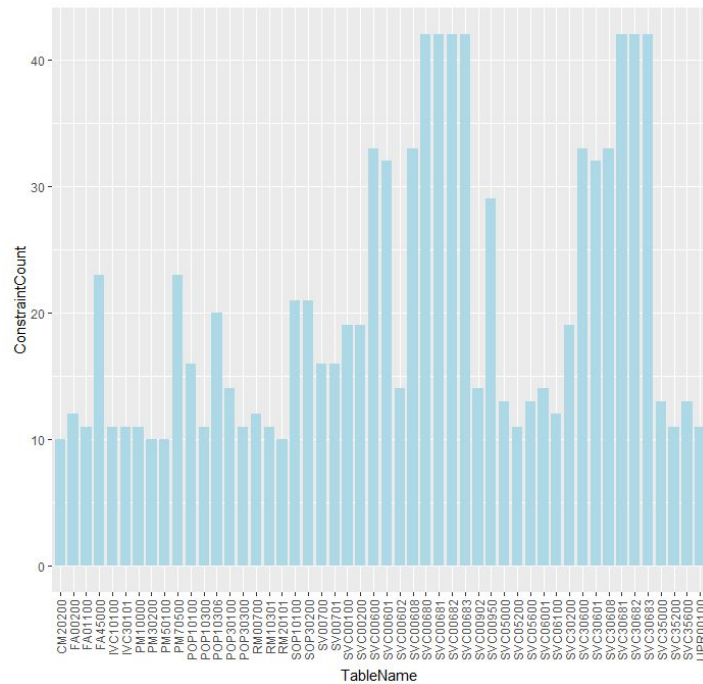
Il faudrait voir la possibilité d'indexer les colonnes les plus demandées, en priorité les tables avec peu de clefs principales. Il va falloir aussi ajouter de clés étrangères pour alléger les vues, procédures et fonctions.

---

Il faudrait commencer par les tables utilisées par le code et par les workflow des scénarios testés en novembre.

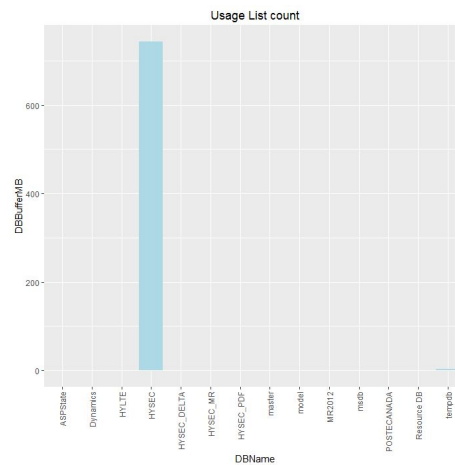
## Annexes

### Figure Analyse de BD-Contraintes



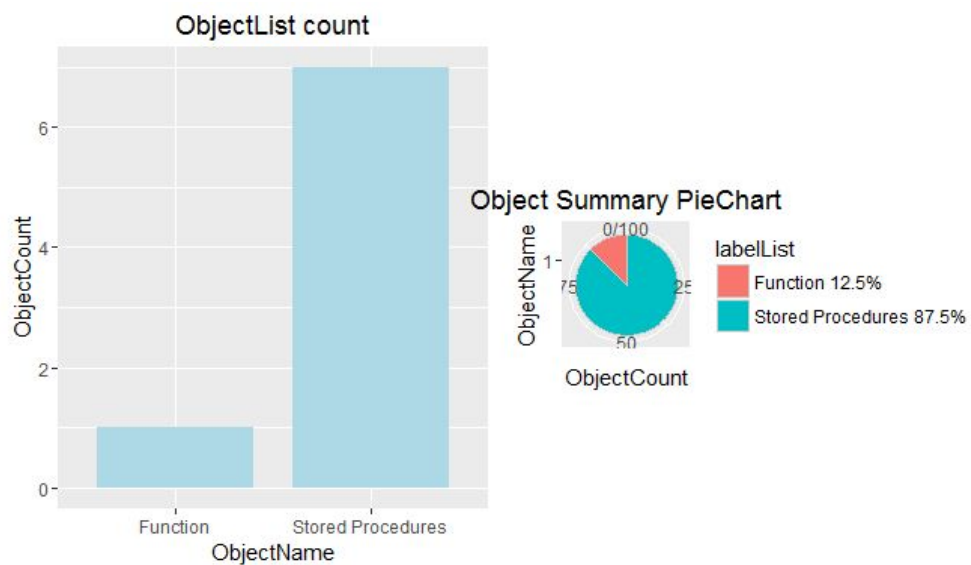
Liste de tables sensible, par table, du ratio nombre de contraintes par table

### Figure Analyse de BD-Mémoire



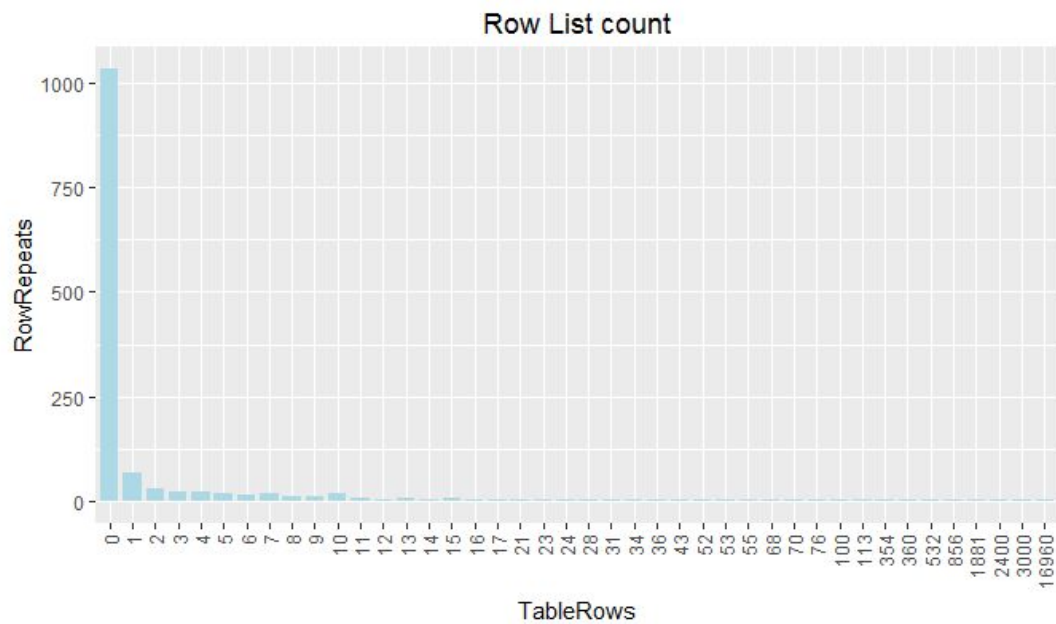
Liste de consommation mémoire par BD

### Figure Analyse de BD-Objets



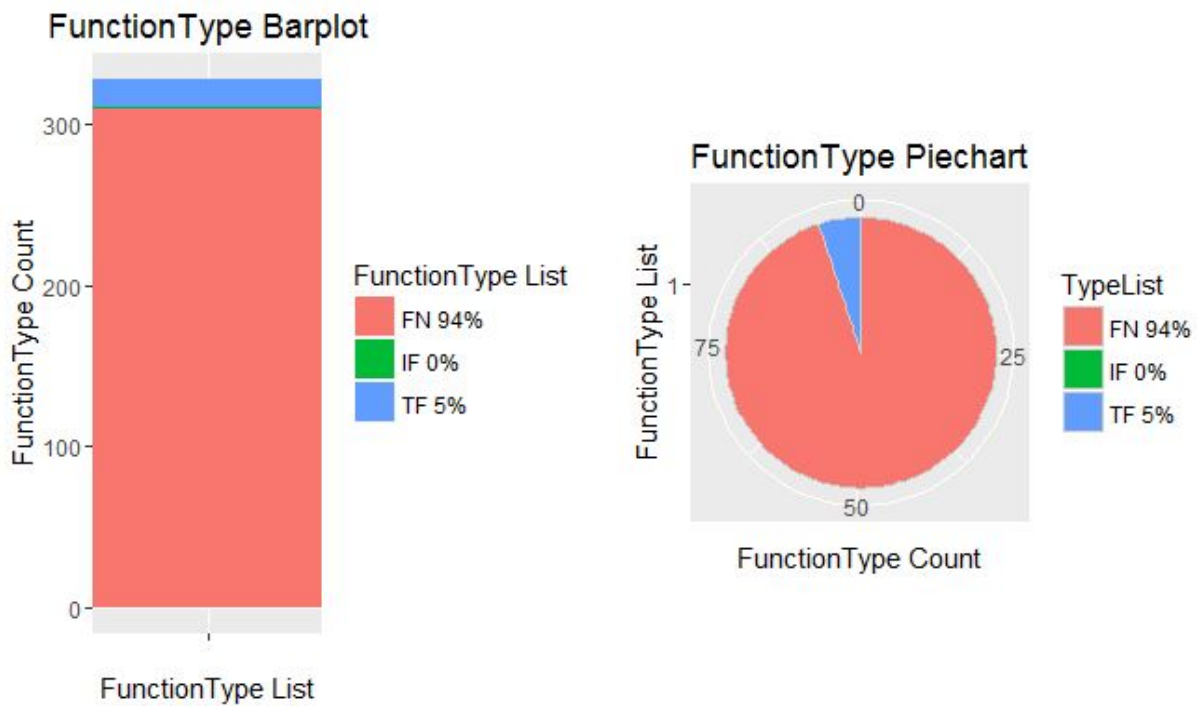
Liste de l'utilisation d'objets, fonctions versus procédures.

### Figure Analyse de BD-Enregistrements de Tables



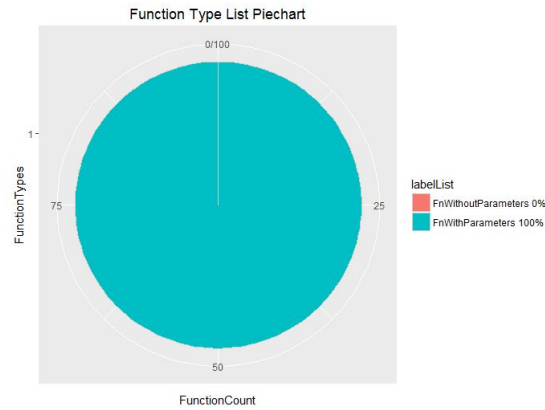
Tri décroissant du nombre d'enregistrements par table

### Figure Analyse de BD-Fonctions



Prédominance de fonctions scalaires

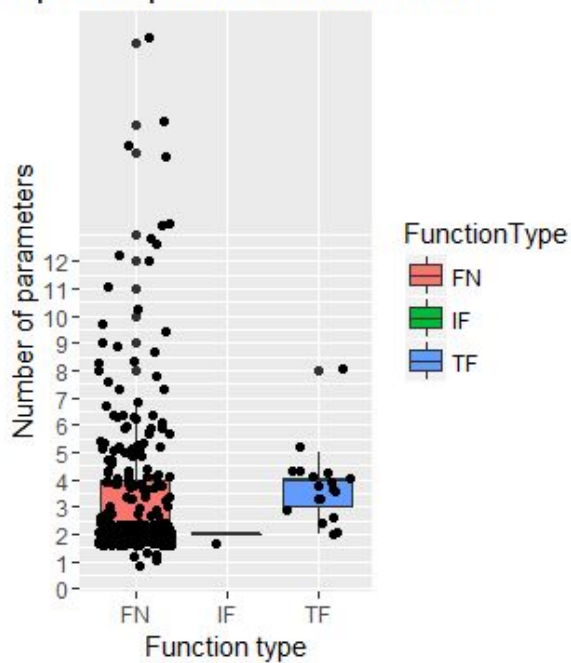
## Figure Analyse de BD-Paramètres de Fonctions



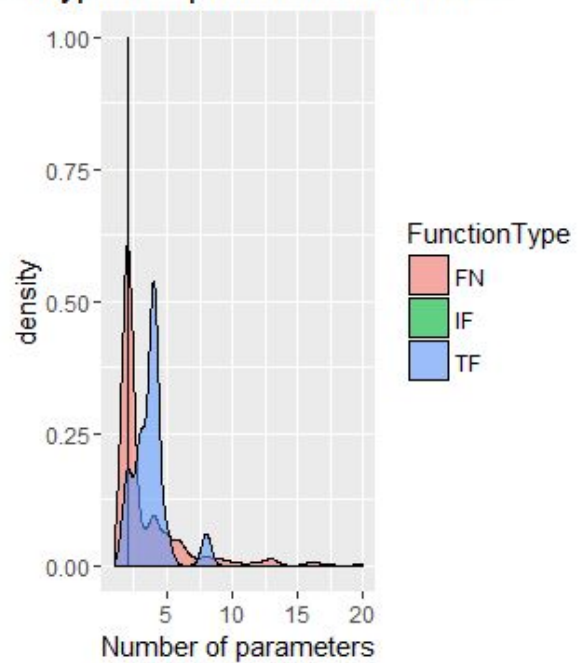
Ratio de fonctions avec paramètres versus sans paramètres

## Figure Analyse de BD-Distribution de paramètres de Fonctions

Boxplot for parameter distribution



Densityplot for parameter distribution

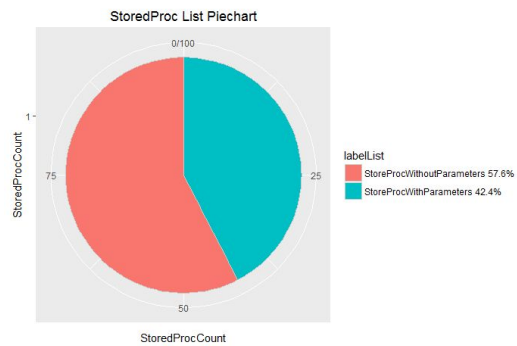


Le nombre de paramètres est dans la norme, peu d'optimisation à faire.



---

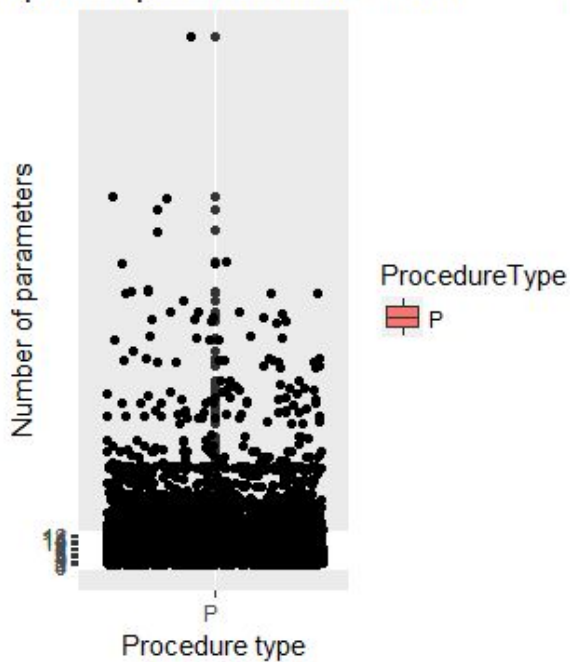
## Figure Analyse de BD-Paramètres de procédures stockées



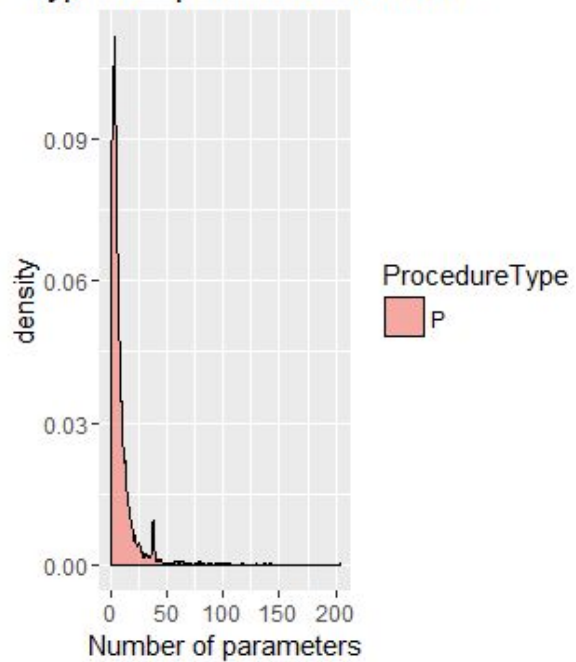
Ratio de 45/55 avec paramètres versus sans paramètres

## Figure Analyse de BD-Distribution de paramètres de procédures stockées

oxplot for parameter distribution

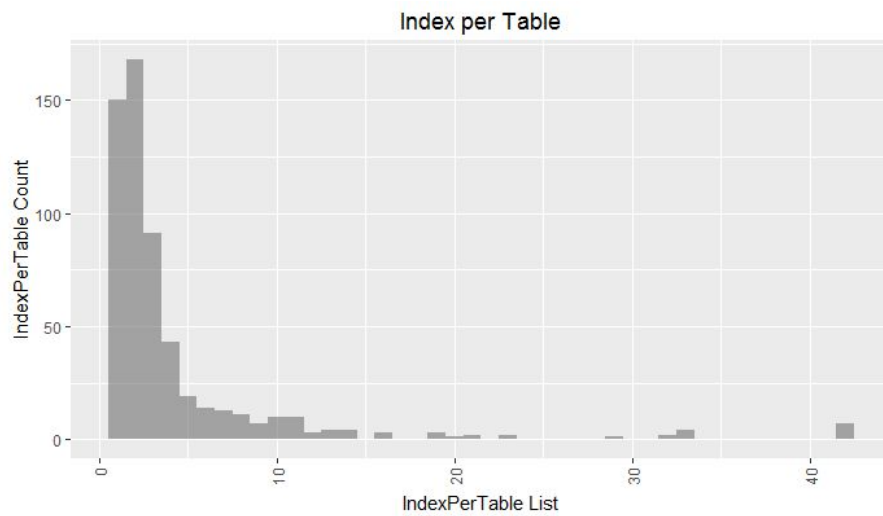


Densityplot for parameter distribution



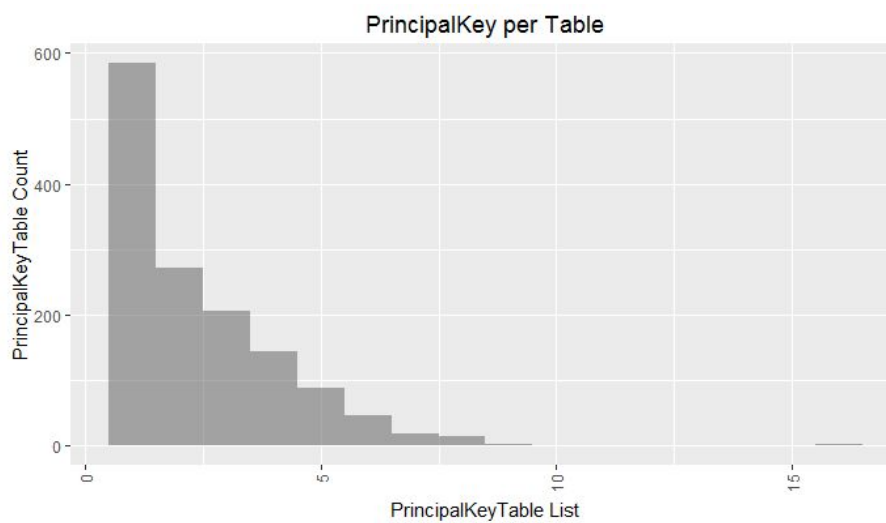
Le nombre de paramètres dépasse 20 avec un pic à 40

## Figure Analyse de BD-Indexes



Histogramme de distribution du nombre d'index par table

## Figure Analyse de BD-Clefs Principales



Nombre de clefs principales par table, 20% ont moins de 3 clefs

## Tableau Analyse de BD-Clefs Étrangères

TableName	ColumnName	ForeignKey	ForeignKeyID	ReferenceTableName	ReferenceColumnName
GMP_AUTO_SCRIPT	ScriptGroupID	FK_ScriptGroup	1513041538	HS_ScriptGroup	ID

---

GMP_ETAT_COMPTE_Detail	SID	FK_GMP_ETAT_CO__SID_5472C43C	1416807484	GMP_ETAT_COMPTE_Sommaire	SID
HS_Client_HS_PortailMessage	HS_PortailMessageID	FK_HS_Client_HS_PortailMessage_HS_PortailMessage	1912809251	HS_PortailMessage	ID
HS_Client_HS_PortailQuestions	HS_PortailQuestionsID	FK_HS_Client_HS_PortailQuestions_HS_PortailQuestions	2008809593	HS_PortailQuestions	ID
HS_COMMUNICATION_USAGERS	UsagerID	FK_HS_COMMUNICATION_USAGERS_USAGERS	347068593	GMP_USAGERS	UsagerID
HS_PortailModifLog	HS_PortailTypeModifID	FK_HS_PortailModifLog_HS_PortailTypeModif	1688808453	HS_PortailTypeModif	ID
RM00105	CPRCSTNM	RM_NationalAccounts_MSTR_FKC	1129053987	RM00101	CUSTNMBR

Notez bien qu'il y en a 7 clés uniquement.