

# La consolidation des données statistiques de Koha dans le réseau BRISE-ES

Symposium Koha - 16 juin 2015

Soledad Beudon - [soledad.beudon@univ-st-etienne.fr](mailto:soledad.beudon@univ-st-etienne.fr)

Université Jean Monnet, Saint-Étienne

BRISE-ES



# Préliminaires...

- ▶ BRISE-ES = Bibliothèques en Réseau Informatisé de Saint-Etienne - Enseignement Supérieur. Réseau de 16 bibliothèques rattachées à 6 établissements différents
- ▶ Début d'utilisation de Koha en production en octobre 2011
- ▶ Koha installé avec SOLR (hors communautaire), actuellement en version 3.06

# À l'origine du projet

Le constat de plusieurs problèmes liés aux statistiques

# Case départ : la table *statistics*

- ▶ La table *statistics* de Koha enregistre les transactions :

datetime	branch	proccode	value	type	other	usercode	itemnumber	itemtype	borrowernumber	associatedborrower
2015-06-01 17:44:54	BUDL	NULL	0.0000	issue		NULL	213214	OUVRAGE	43796	NULL
2015-06-01 17:44:44	BUDL	NULL	0.0000	issue		NULL	556699	OUVRAGE	43796	NULL
2015-06-01 17:44:15	BUDL	NULL	0.0000	return		NULL	479423	EXCLU	43796	NULL
2015-06-01 17:40:17	BUDL	NULL	0.0000	issue		NULL	297350	OUVRAGE	38278	NULL
2015-06-01 17:39:44	ENSASE	NULL	0.0000	issue		NULL	72844	OUVRAGE	51425	NULL
2015-06-01 17:39:39	ENSASE	NULL	0.0000	issue		NULL	4235	OUVRAGE	51425	NULL
2015-06-01 17:39:34	ENSASE	NULL	0.0000	issue		NULL	197541	OUVRAGE	51425	NULL
2015-06-01 17:39:16	BUDL	NULL	0.0000	return		NULL	479434	EXCLU	55440	NULL
2015-06-01 17:38:37	ENSASE	NULL	0.0000	return		NULL	197541	OUVRAGE	51425	NULL
2015-06-01 17:38:32	ENSASE	NULL	0.0000	return		NULL	72844	OUVRAGE	51425	NULL
2015-06-01 17:38:25	ENSASE	NULL	0.0000	return		NULL	4235	OUVRAGE	51419	NULL
2015-06-01 17:34:26	BUDL	NULL	0.0000	return		NULL	479391	EXCLU	34842	NULL
2015-06-01 17:34:00	BUDL	NULL	0.0000	issue		NULL	100604	OUVRAGE	54702	NULL
2015-06-01 17:33:45	BUDL	NULL	0.0000	return		NULL	505672	OUVRAGE	NULL	NULL
2015-06-01 17:32:45	BUDL	NULL	0.0000	return		NULL	479424	EXCLU	44000	NULL
2015-06-01 17:32:15	BUROANNE	NULL	0.0000	renew		NULL	379874	OUVRAGE	61038	NULL

- ▶ Cette table nous a vite posé plusieurs problèmes...

# Non-conformité CNIL

- ▶ La table *statistics* stocke l'identifiant lecteur ET l'identifiant exemplaire pour chaque transaction :

datetime	branch	proccode	value	type	other	usercode	itemnumber	itemtype	borrowernumber	associatedborrower
2015-06-01 17:44:54	BUDL	NULL	0.0000	issue		NULL	213214	OUVRAGE	43796	NULL
2015-06-01 17:44:44	BUDL	NULL	0.0000	issue		NULL	556699	OUVRAGE	43796	NULL

- ▶ À l'origine, sur notre installation, pas de nettoyage spécifique prévu pour cette table
- ▶ Si on la laisse telle quelle, on peut reconstituer tout l'historique de prêt d'un lecteur sans limitation de durée
- ▶ Or la réglementation CNIL limite à 4 mois la conservation de telles données une fois la transaction terminée (<http://www.cnil.fr/vos-obligations/declarer-a-la-cnil/declaration/mon-secteur-dactivite/mon-theme/je-dois-declarer/declaration-selectionnee/dec-mode/DISPLAYSINGLEFICHEDECL/dec-uid/17/> )

# Écart entre l'enregistrement et l'extraction (1/2)

- ▶ En dehors des identifiants lecteur / exemplaire, très peu de données enregistrées dans la table *statistics* au fil de l'eau
- ▶ Pour croiser les données des transactions avec d'autres données lecteurs (la catégorie, le cursus...) ou exemplaires (la cote, la localisation...), il faut remonter aux tables lecteurs ou exemplaires à travers des jointures
- ▶ **Les données obtenues par ces jointures correspondent au moment où on fait l'extraction et non au moment d'enregistrement de la transaction**
- ▶ **Plus le moment de l'extraction est éloigné du moment de l'enregistrement de la transaction, plus le risque de croiser des données modifiées au fil du temps, et donc d'obtenir des résultats faux, augmente**

# Écarts entre l'enregistrement et l'extraction (2/2) : Cas fréquents

Au niveau des exemplaires :

- ▶ Changements de cote
- ▶ Changements de localisation : libre accès > magasin, déménagement...
- ▶ Désherbage, suppression d'exemplaires

Au niveau des dossiers lecteurs :

- ▶ Changement d'année d'études
- ▶ Changement de catégorie
- ▶ Changement de cursus
- ▶ Suppression du dossier

# Au niveau local, deux besoins supplémentaires

- ▶ Avoir un rétrospectif sur plusieurs années
  - ▶ Mise en place d'une politique documentaire
  - ▶ Volonté de récupérer la table statistique de notre SIGB précédent (Loris)
- ▶ Pouvoir interroger une base à jour sans peser sur la base de production

# Une solution stéphanoise :

Mise en place de tables statistiques spécifiques sur  
une base Koha à j-1

# Mise en place de l'import journalier de la base de production

- ▶ Sauvegarde des permissions (champ flags), des rapports sauvegardés (table saved\_sql) et des préférences système de la base j-1
- ▶ Import de la base de données de production : cet import écrasera les tables Koha mais pas nos tables spécifiques
- ▶ Import des index SOLR afin de ne pas devoir lancer l'indexation complète de la base j-1
- ▶ Réintroduction des permissions, des rapports sauvegardés et des préférences système de la base j-1

# Constitution de tables spécifiques (1/2)

Mise en place de 5 tables statistiques brisees\_ :

- ▶ Emprunts (prêts et prolongations)
- ▶ Réservations (réservations finies)
- ▶ Acquisitions (lignes de commande reçues)
- ▶ Fascicules (fascicules reçus)
- ▶ Lecteurs actifs (table supprimée ultérieurement)

# Constitution de tables spécifiques (2/2)

- ▶ Récupération des données de la table statistique de notre ancien SIGB Loris (2003-2011)
- ▶ Récupération des données du rétrospectif Koha
- ▶ Mise en place de l'alimentation quotidienne des tables spécifiques dans la foulée de l'import quotidien de la base de production en base j-1 :
  - ▶ Les valeurs Koha sont importées sous forme de code
  - ▶ La valeur SQL NULL est remplacée lors de l'import par des vides ou des 0
  - ▶ On n'importe aucun champ multivalué

# Exemple : l'alimentation de la table brisees\_emprunts

- ▶ Insertion des données transactions (statistics)
- ▶ Mise à jour des données :
  - ▶ exemplaires (items)
  - ▶ lecteurs (borrowers)
  - ▶ attributs lecteurs (borrower\_attributes)
  - ▶ biblio 1 (biblioitems)
  - ▶ biblio 2 (biblio)
- ▶ Anonymisation des champs datetime\_tempo et borrowernumber\_tempo



Brisees\_emprunts



Requetes\_emprunts

# Les évolutions ultérieures

- ▶ Cryptage de l'identifiant lecteur dans `brisees_emprunts` et `brisees_réservations`
- ▶ Suppression de la table `brisees_lecteurs_actifs`
- ▶ Mise à jour rétrospective des domaines des exemplaires dans `brisees_emprunts` et `brisees_réservations`
- ▶ À chaque début d'année, suppression des données de l'année la plus ancienne afin de garder 10 ans de statistiques courantes
- ▶ Augmentation de l'espace disque (la table `brisees_emprunts` contient 1,8 M d'enregistrements...)

# Résultat : un bilan très positif

- ▶ Gain de fiabilité : données homogènes enregistrées au fil de l'eau
- ▶ Résultats statistiques enrichis sur les emprunts, les réservations, les acquisitions et les fascicules
  - ▶ Ex : emprunts par support, par type de carte, domaine...



Transactions2014



TransactionsDomaines

- ▶ On ne rencontre plus de problèmes liés à des suppression d'exemplaires, de dossiers lecteurs ou d'abonnements
- ▶ On a pu répondre à l'enquête ESGBU portant sur les lecteurs actifs 2012-2013 et 2013-2014 mais parue seulement au printemps 2015.
- ▶ On a pu mettre en place en production l'anonymisation des transactions, dont la table *statistics* et le nettoyage des dossiers lecteurs pour conformité CNIL

# Exploitation du module « Bilans et statistiques »

Modifications et adaptations locales

# L'idée initiale...

- ▶ Ne pas dépendre des « Assistants statistiques » de Koha peu adaptés à nos besoins
- ▶ Construire une bibliothèque de rapports sauvegardés limitée en nombre mais souple en usages
- ▶ Rendre l'utilisation des rapports aisée et lisible par un groupe d'utilisateurs non experts en MySQL
- ▶ Leur laisser la plus grande marge de manœuvre et de choix en privilégiant les rapports à variable

# Amélioration du tableau des rapports sauvegardés

On a ajouté le plugin DataTables de jquery au tableau qui contient les rapports sauvegardés, ce qui nous permet :

- ▶ De trier les colonnes

ID	Nom du rapport	Type	Notes
----	----------------	------	-------

- ▶ De choisir le nombre de rapports affichés

## Rapports sauvegardés

Choisir le rapport à exécuter à partir de la liste

Montrer  entrées par page

- ▶ De filtrer le contenu de tous les champs du tableau

filtrer (sur toutes les colonnes et toutes les lignes) :

# Les rapports sauvegardés

- ▶ Rapports préfixés en fonction du module avec couleur spécifique  
ACQ -, CAT-, CIRC -, PERIO - pour le tri :

ID	Nom du rapport	Type	Notes
41	<b>ACQ - Décompte</b> des dépenses à un instant t avec comparaison calendrier	1	Décompte des dépenses à un instant t par poste budgétaire avec comparaison avec l'estimé du calendrier et écart estime-dépense Variable : poste budgétaire Tables Koha acq
2	<b>ACQ - Décompte</b> des exe reçus par poste budgétaire avec détails	1	Décompte des exe reçus et sommes des montants dépensés, répartis par poste budgétaire, fournisseur, domaine, provenance, code de collection Variable : poste budgétaire Table brisees_reception_acq

- ▶ Rapports avec deux types de variables :

- ▶ Menu déroulant avec liste de valeurs autorisées

Ex : items.homebranch=<<Bibliothèque de l'exemplaire | branches>>

- ▶ Champ libre

Ex : items.itemcallnumber like << Cote commence par (mettre % derrière début de cote ou % seul pour toutes les cotes) >>

# Au lancement des rapports

- ▶ Les rapports avec variables ne se lancent pas immédiatement quand on clique sur « exécuter », mais ouvrent une page « Saisir les paramètres... »
- ▶ On a modifié `guided_reports.pl` pour reporter sur cette page le champ « notes » du tableau : l'utilisateur a ainsi sous les yeux la description du rapport au moment de le lancer
  - ▶ Exemple de rapport CAT - prêt à l'emploi :

Saisir les paramètres pour le rapport **CAT - Liste** d'exe par bib, loc, cote, date de création de l'exemplaire et code de collection biblio:

Liste d'exe par bibliothèque, localisation, cote, date de création de l'exe et code de collection biblio

Variables: bibliothèque, localisation, cote, dates de création de l'exe et code de collection biblio

Champs exportés : Code de collection notice, Date de création Notice, Auteur, Titre, Complément de titre, PPN, ISBN, ISSN, Editeur, Date de publication, Titre collection, Volume collection, Bibliothèque, Localisation, Code barres, Cote, Numéro inventaire, Date de création exemplaire, imputation, Notes exemplaire, Numero de périodique, Nombre de prêts, Date de retour prévue

Tables Koha biblio et exe

Code de collection de la notice :

ET Bibliothèque de l'exemplaire :

ET localisation de l'exemplaire contient (mettre la loc sous forme de code ou % seul pour toutes les loc) :

ET Cote commence par (mettre % derrière début de cote ou % seul pour toutes les cotes) :

ET Date de création de l'exemplaire inférieure ou égale à (mettre sous la forme AAAA-MM-JJ) :

ET Date de création de l'exemplaire supérieure ou égale à (mettre sous la forme AAAA-MM-JJ) :

Exécuter le rapport

# Deux petites astuces SQL pour faciliter la lecture des résultats

- ▶ IFNULL pour distinguer les champs non renseignés des champs NULL

GATELSE	BIB	OUVRAGE	LIVRE	GATE-ECONOMIE	0	34200012506685	A12 REI	GAT 8079	2013-11-21	NULL	
---------	-----	---------	-------	---------------	---	----------------	------------	----------	------------	------	--

- ▶ WITH ROLLUP (éventuellement avec CASE) pour afficher des totaux, sous-totaux

## ACQ - Décompte des exe reçus par poste budgétaire global

Décompte des exe reçus et sommes des montants dépensés, répartis par budget père, poste budgétaire

Variable : poste budgétaire

Table brisees\_reception\_acq

Nombre total d'enregistrement correspondant à la requête : 6.

```
SELECT ifnull(budget_pere, 'Total') as budget_pere, case WHEN poste_budgetaire is not null THEN poste_budgetaire WHEN budget_pere IS NOT NULL THEN 'Sous-total' ELSE sum( nbr_exe_recus ) AS 'quantite recue' FROM brisees_reception_acq WHERE 'poste_budgetaire' LIKE '%BUDL%2014%' GROUP BY budget_pere, 'poste_budgetaire' with
```

budget_pere	poste_budgetaire	somme prix fois quantite	quantite recue
BUDL CQ 2014	BUDL CQ 2014	27975,87	1316
BUDL CQ 2014	Sous-total	27975,87	1316
BUDL DGF 2014	BUDL DGF PEDA 2014	31258,64	1244
BUDL DGF 2014	BUDL DGF RECH 2014	6692,58	219
BUDL DGF 2014	Sous-total	37951,22	1463
Total	Total	65927,09	2779

Télécharger le rapport

Merci de votre attention