

Université IBM i

10 et 11 mai 2016 – IBM Client Center de Bois-Colombes

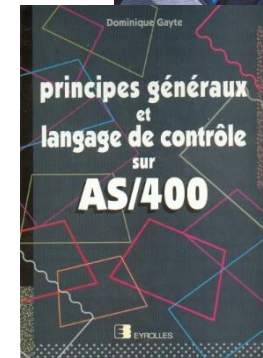
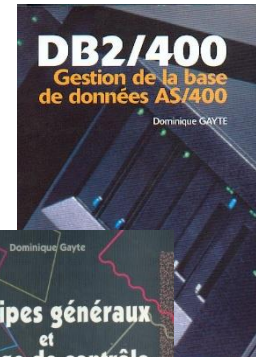
S28 – Analytique : pourquoi et comment construire un datawarehouse avec l'ETL de DB2 for i ?

Mercredi 11 mai – 11h00-12h30

Dominique GAYTE – NoToS
dgayte@notos.fr – www.notos.fr

NoToS

- Expertise autour de l'IBM i
 - Sécurité
 - Regard moderne (DB2 Web Query)
- Service
 - Formation, audit, développement...
- PHP sur IBM i avec Zend
- Développement de progiciels
 - Modernisation à valeur ajoutée des IBM i



Valorisation des spools des IBM i (AS/400)
Transformation en PDF, archivage, indexation
<http://www.notos.fr/phpSpool.aspx>



Gestion de Contenu (ECM)
GED, graphiques, alertes, workflow, GANTT...
<http://www.lorena.pro>

Sommaire

- La problématique
- Qu'est ce qu'un Datawarehouse
- Un Datawarehouse sur l'IBM i ?
- Créer et alimenter un Datawarehouse
- Data Migrator for i



La problématique

Structure de la base de données

- Les données de production ne sont pas organisées pour répondre aux besoins de reporting/décisionnel
 - Beaucoup
 - D'enregistrements
 - De jointures
 - De zones
 - Données élémentaires, non agrégées
 - Détail des factures
 - Dates sous des formats divers
 - Traitements à effectuer
 - Une zone contient plusieurs informations
 - Sélections sur les lignes
 - Mouvements de stocks dans les ventes

Temps de réponses

- Pour optimiser les temps de réponse, on aurait intérêt à :
 - Faire des agrégations
 - CA mensuels par produit, commercial, par client...
 - Récapitulatifs par mois, par agence
 - Stocker le résultat de calculs
 - Rajouter des informations
 - Historique
 - Altérer la troisième forme normale
- Utiliser des structures adaptées
 - Tables matérialisées (MQT)
 - Index

Hétérogénéité des données

- Les données utiles sont dans de nombreux environnements
 - Bases de données
 - DB2 for i (diverses partitions)
 - SQL Server, Oracle, MySQL
 - ...
 - Fichiers plats
 - Excel
 - XML, TXT...

Temps réel ?

- Les données de la BD de production sont maintenues en temps réel
- En reporting/décisionnel on a rarement besoin de disposer de données instantanées
 - On préfère privilégier les performances
- On peut donc utiliser une base de données optimisée pour le reporting/décisionnel
 - Mise à jour régulièrement
- Les requêtes nécessitant des données en temps réel sont exécutées sur la BD de production

Qu'est ce qu'un Datawarehouse ?

- Les entrepôts de données sont des bases de données organisées pour les besoins du décisionnel
- Eventuellement composé de datamarts (magasins de données) chacun étant orienté vers un métier
- Quelques caractéristiques des DW
 - Forte volumétrie
 - Utilisé en lecture
 - Surtout avec SQL
 - Forte Sécurité
 - Rapidité d'accès



Un Datawarehouse sur l'IBM i ?

Datawarehouse et IBM i

- Rappel : Un Datawarehouse est une base de données (optimisée pour le reporting/décisionnel)
- Rappel 2 : DB2 for i
 - Est un des meilleurs SGBD du marché
 - Répond aux contraintes énoncées (sécurité, SQL, volumétrie...)
- L'IBM i peut très bien héberger un Datawarehouse
- C'est même une excellente solution !
- Comment ?
 - Dans une ou plusieurs bibliothèques
 - Dans une partition dédiée
 - CPU, mémoire, disques dédiés

Créer et alimenter un Datawarehouse

Création et alimentation

■ Création

- Il faut préférer SQL plutôt que les DDS
- Utiliser toutes les possibilités de DB2 for i
 - Journalisation (en cas de mise à jour erronée)
 - Les vues, les index
 - Les fonctions
 - Les tables matérialisées
 - Les outils de gestion des performances
 - Les procédures stockées, les Web Services

■ Alimentation

- Ils sont alimentés par des données de production
- Le processus consiste à extraire les données de production, à les traiter (agrégations, calculs...) et à les charger dans le datawarehouse
- Les ETL (Extract, Transform & Load) sont des outils dédiés

Mécanismes d'alimentation

- En CL
- En RPG/COBOL
- Avec un ETL
 - Data Migrator for i

Alimenter un DW avec un CL

- Commandes de l'IBM i
 - CPYF
 - CPYFRMIMPF

- SQL
 - RUNSQLSTM
 - RUNSQL (V7R1) utilisant les variables du programme
 - Rafraichissement des MQT

Alimenter un DW en RPG

- Natif pour DB2 for i en local
- En RPG IV (et SQL) on sait
 - Lire les bases externes (JDBC)
 - SQL Server, Oracle, MySQL...
 - Lire et parser des fichiers XML
 - Traiter les données et les écrire dans le DW
- Demande un important travail de codification
- N'est acceptable que pour des DW peu complexes

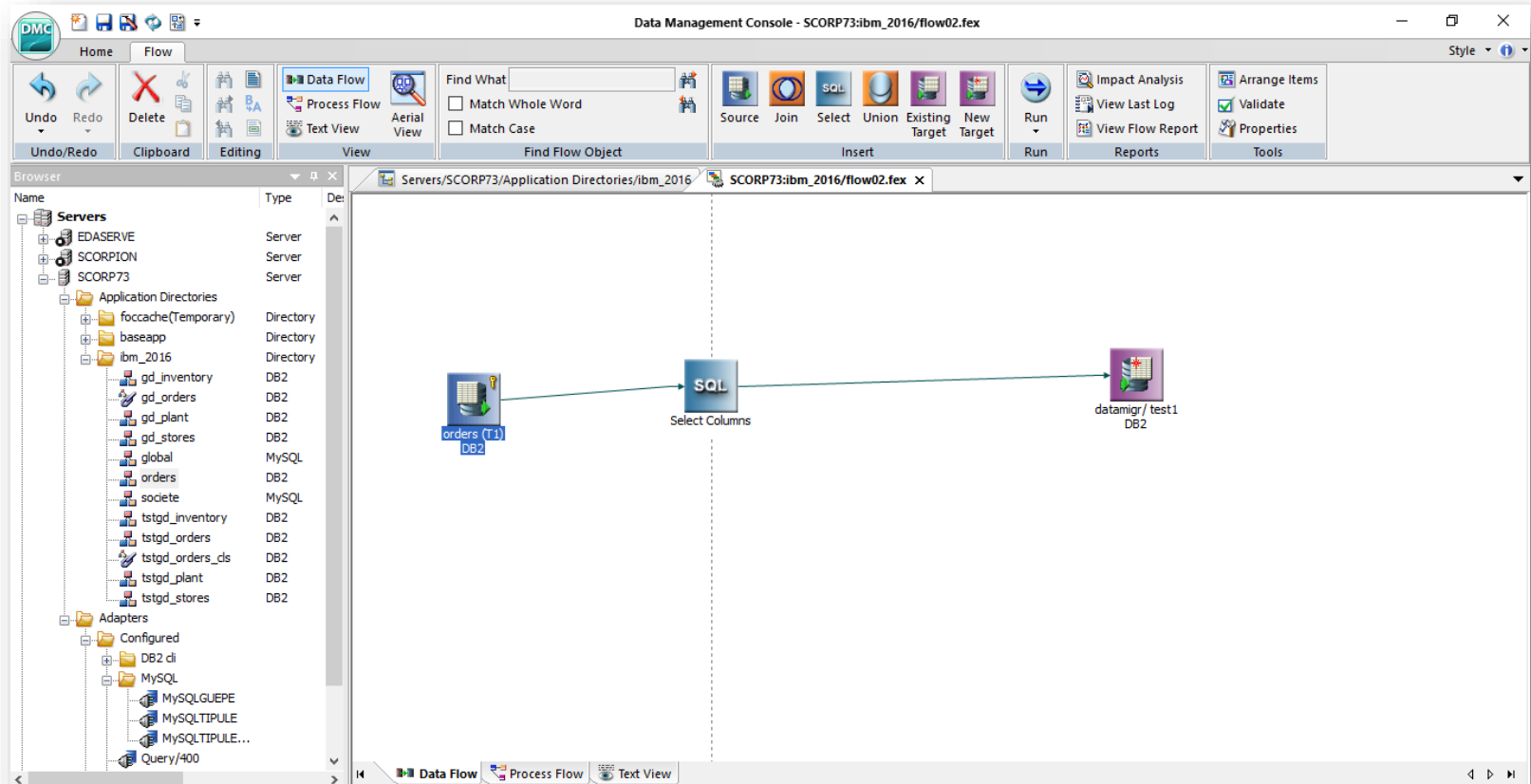
Les ETL

- Outils conçus pour alimenter les DW
- Il existe des grands standards sur le marchés
 - Souvent très chers
 - L'IBM i est souvent vue comme un SGBD marginal
- Data Migrator for i est l'ETL associé à DB2 Web Query
 - Natif IBM i

Data Migrator for i

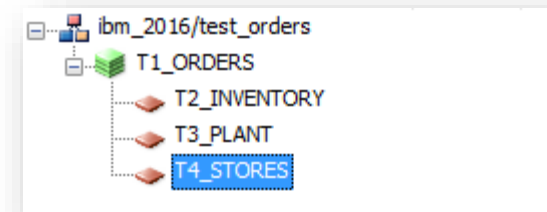
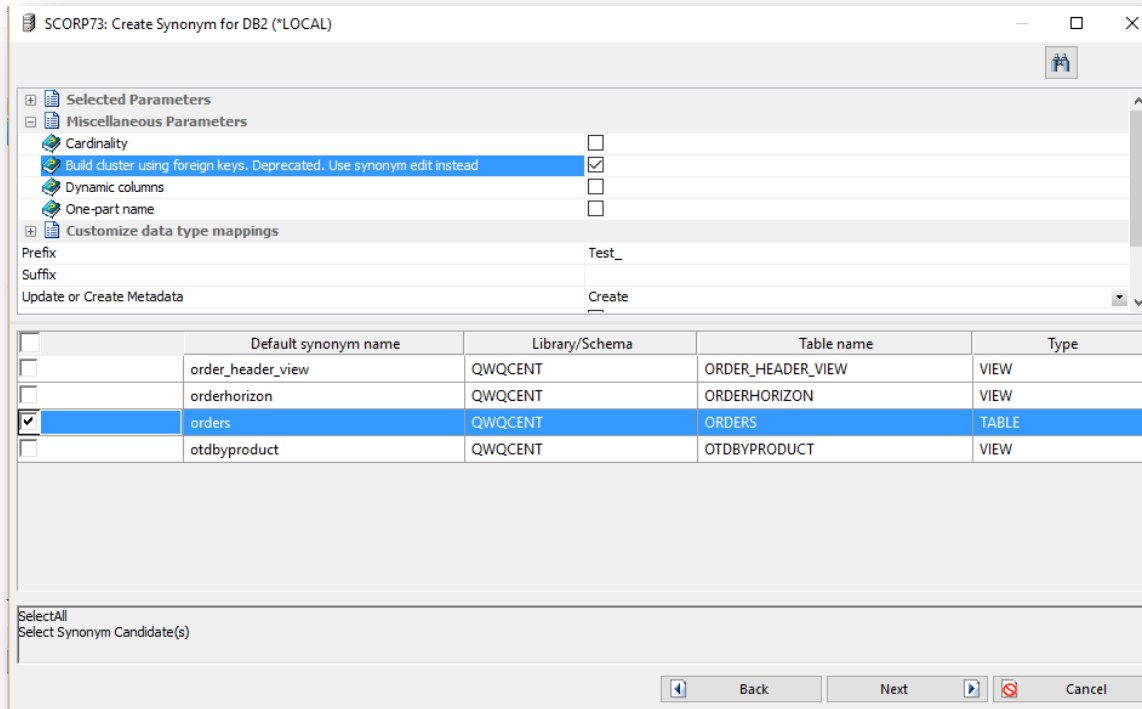
Data Migrator for i

- Logiciel facturable, option de DB2 Web Query
- Interface graphique basée sur un Studio
 - Data Management Console



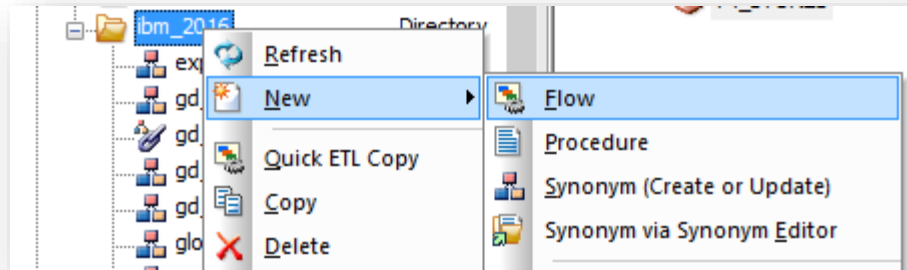
Création du premier flux de données

- Utilisation du schéma (bibliothèque !) QWQCENT livré avec DB2 Web Query
- Création d'une métadonnée sur le table des commandes (ORDERS) avec intégration automatique des CIR
 - Sera utilisée comme source

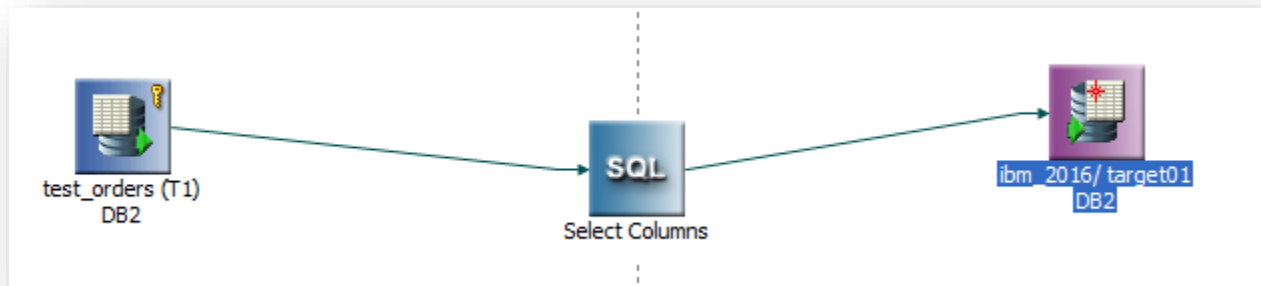


Création du premier flux de données (2)

- Nouveau flux



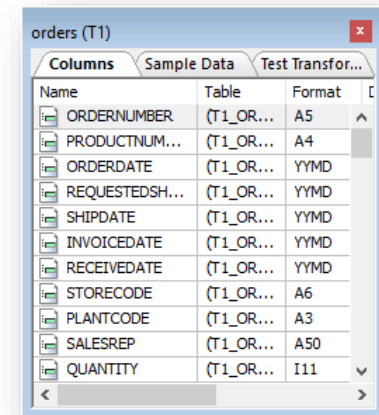
- Sélection de la source, des traitements et de la cible



Source



- A partir d'une métadonnée existante
- Par glisser déplacer dans la partie gauche de la fenêtre
- Sample Data : échantillon de données
- Toogle : visualisation des colonnes
- Data profiling
 - Vue statistique

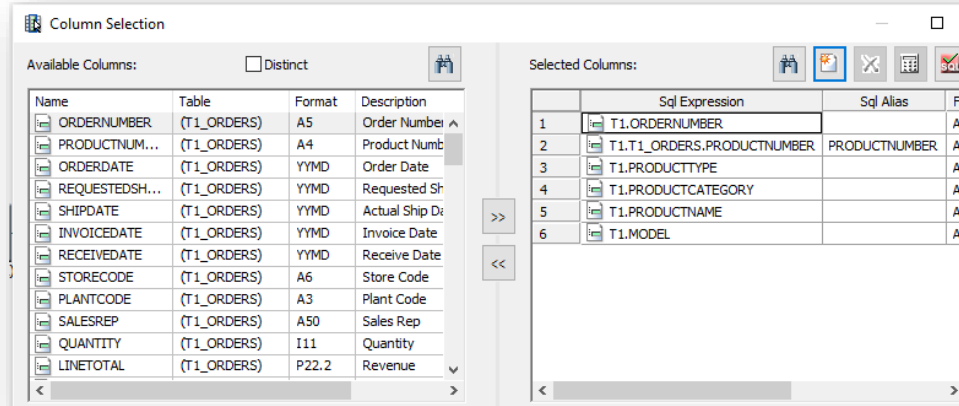


	Name	Format	Count	Distinct Count	Distinct Percent	Patterns Count	Minimum Value	Maximum Value	Average Value	Minimum Length	Maximum Length	Average Length	Nulls Count	Nulls Percent	Duplicate Count	Duplicate Percent
1	ORDERNUMBER	A5	32283	5783	17.91	<u>2</u>	28003	9955A	.	5	5	5	0	0.00	26500	82.09
2	PRODUCTNUMBER	A4	32283	75	0.23	<u>1</u>	1001	5015	.	4	4	4	0	0.00	32208	99.77
3	ORDERDATE	YYMD	32283	475	1.47	.	2012/01/01	2013/12/29	2012/12/09	.	.	.	0	0.00	31808	98.53
4	REQUESTEDSHIPDATE	YYMD	32283	788	2.44	.	2012/02/01	2014/04/05	2013/02/06	.	.	.	0	0.00	31495	97.56
5	SHIPDATE	YYMD	32283	825	2.56	.	2012/01/27	2014/06/13	2013/02/12	.	.	.	0	0.00	31458	97.44
6	INVOICEDATE	YYMD	32283	828	2.56	.	2012/02/03	2014/06/23	2013/02/21	.	.	.	0	0.00	31455	97.44
7	RECEIVEDATE	YYMD	32283	834	2.58	.	2012/02/06	2014/07/07	2013/02/28	.	.	.	0	0.00	31449	97.42

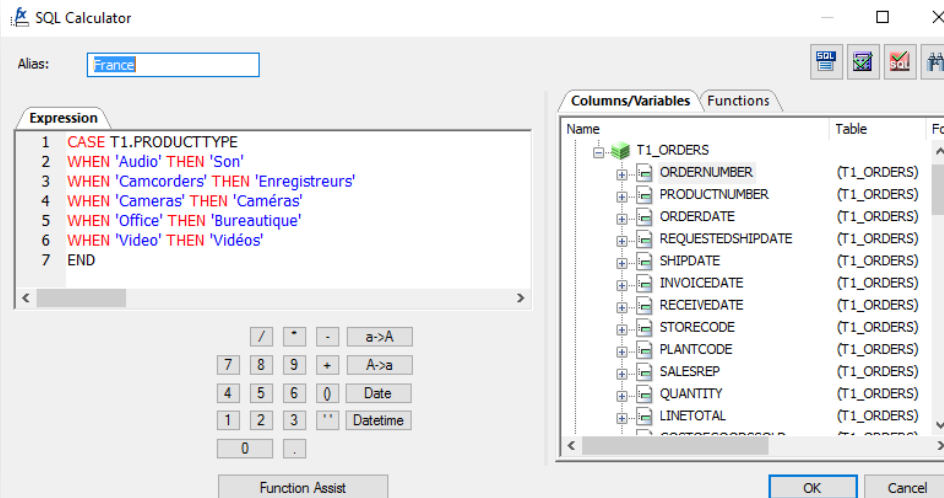
Transformation



■ Sélection des colonnes



■ Ajout de colonnes

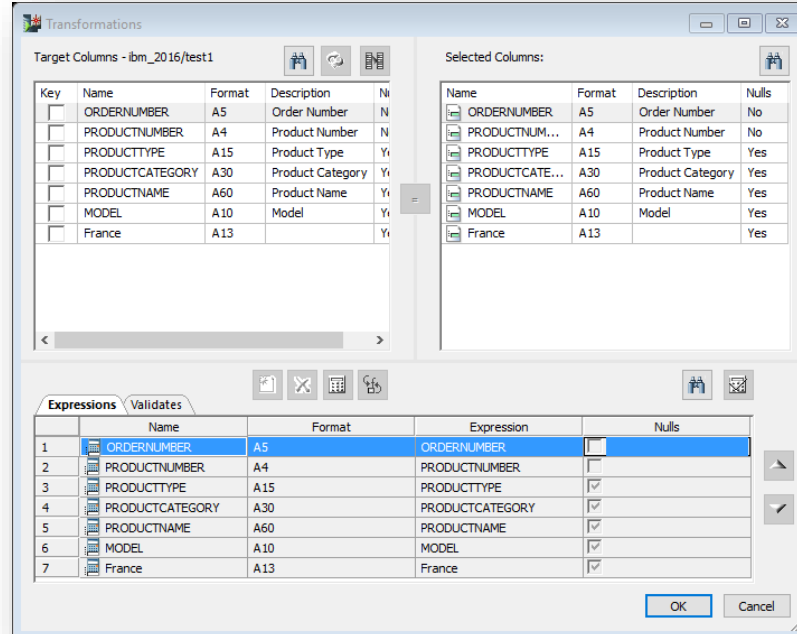




Cible

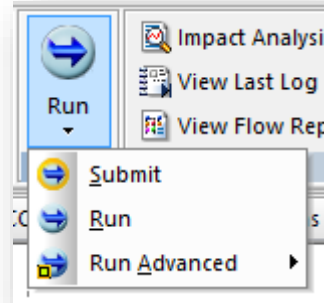


- Par glisser déplacer dans la partie droite
- Définition des clés possible
- Dans les propriétés
 - Choix de la connexion
 - Table cible si c'est une nouvelle table
 - Création du synonyme















Exécution (1)

- A la demande
 - Batch
 - Interactif

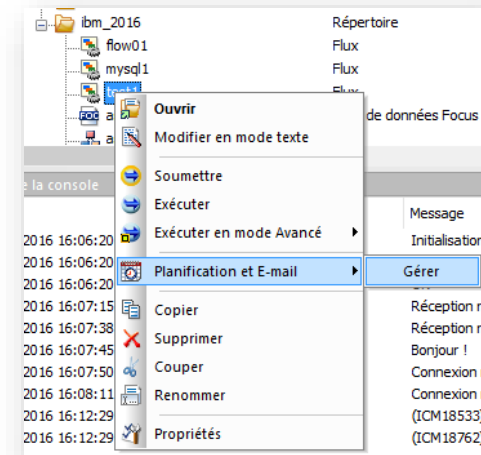


- Console Log pour visualiser ce qui s'est passé

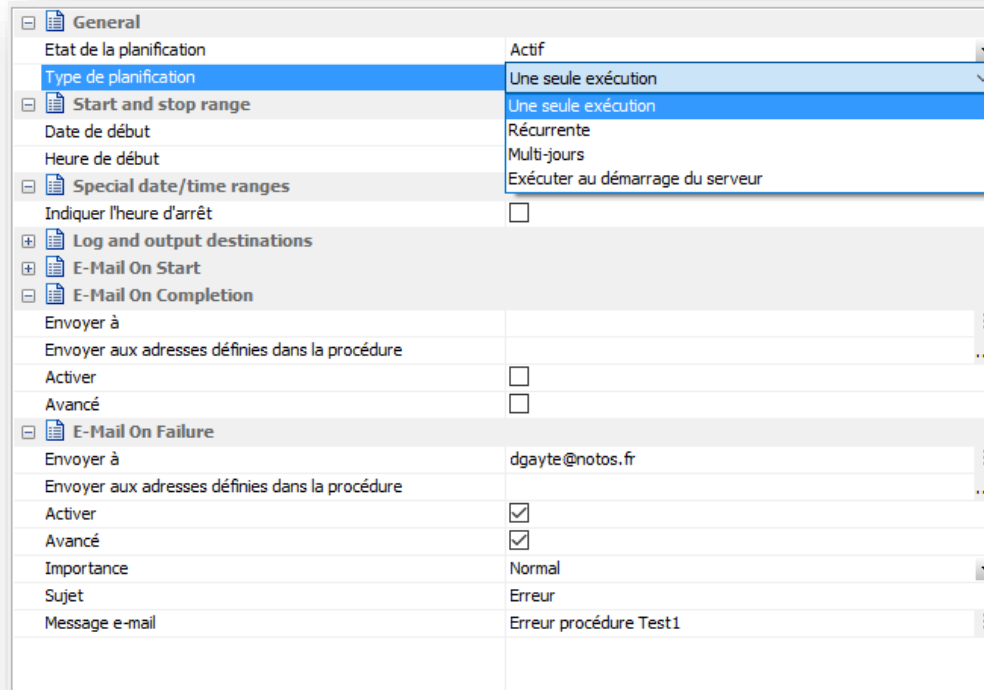
Console Log		
Time	Server	Message
 05/02/2016 15:43:24	SCORP73	(ICM18514) 32283 : Row(s) accepted by target table
 05/02/2016 15:43:24	SCORP73	(ICM18515) 32283 : Row(s) inserted into target table
 05/02/2016 15:43:24	SCORP73	(ICM18517) 0 : Row(s) updated in target table
 05/02/2016 15:43:24	SCORP73	(ICM18518) 0 : Row(s) deleted from target table
 05/02/2016 15:43:24	SCORP73	(ICM18520) 0 : Row(s) rejected due to validation errors
 05/02/2016 15:43:24	SCORP73	(ICM18521) 0 : Row(s) rejected due to no match
 05/02/2016 15:43:24	SCORP73	(ICM18522) 0 : Row(s) rejected because duplicate exist
 05/02/2016 15:43:24	SCORP73	(ICM18808) 0 : Row(s) rejected due to DBMS error
 05/02/2016 15:43:24	SCORP73	(ICM18744) Ending Load
 05/02/2016 15:43:24	SCORP73	(ICM18040) Return Code = 0
 05/02/2016 15:43:24	SCORP73	(ICM18076) Request: __WCFEX - finished processing
 05/02/2016 15:43:24	SCORP73	(ICM18007) CPU Time : 1535

Exécution (2)

- Planification

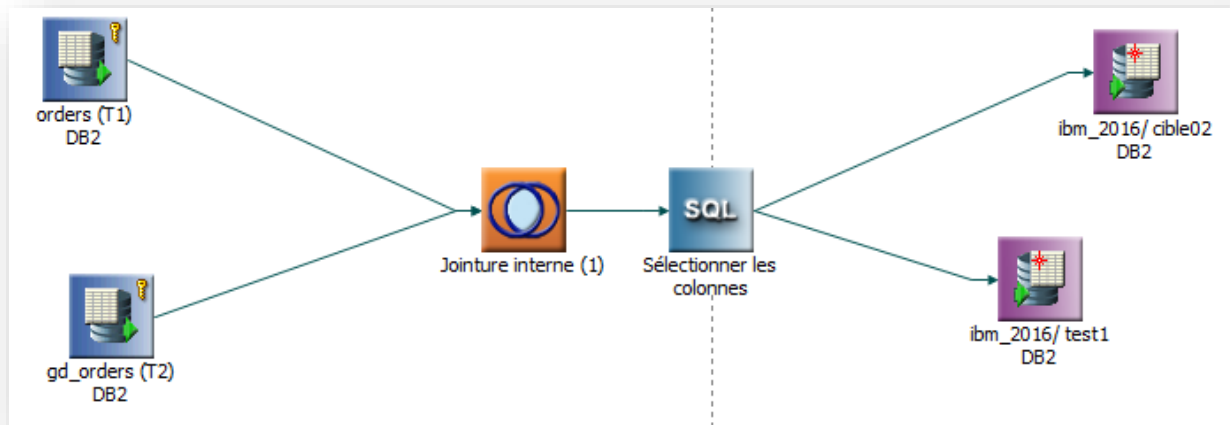


- Choix des évènements associés aux e-mail



Sources : jointures

- Jointure avec assistant



Editeur de jointures

Source gauche :			Source droite :		
Nom	Table	Format	Nom	Table	Format
ORDERNUMBER	(T1_ORDERS)	A5	ORDERNUMBER	(GD_ORDERS)	A5
PRODUCTNUMBER	(T1_ORDERS)	A4	PRODUCTNUMBER	(GD_ORDERS)	A4
ORDERDATE	(T1_ORDERS)	YYI	ORDERDATE	(GD_ORDERS)	YYI
REQUESTEDSHIPDATE	(T1_ORDERS)	YYI	REQUESTEDSHIPDATE	(GD_ORDERS)	YYI
SHIPDATE	(T1_ORDERS)	YYI	SHIPDATE	(GD_ORDERS)	YYI
INVOICEDATE	(T1_ORDERS)	YYI	INVOICEDATE	(GD_ORDERS)	YYI
RECEIVEDATE	(T1_ORDERS)	YYI	RECEIVEDATE	(GD_ORDERS)	YYI
STORECODE	(T1_ORDERS)	A6	STORECODE	(GD_ORDERS)	A6
PLANTCODE	(T1_ORDERS)	A3	PLANTCODE	(GD_ORDERS)	A3

Type de jointure : Jointure interne

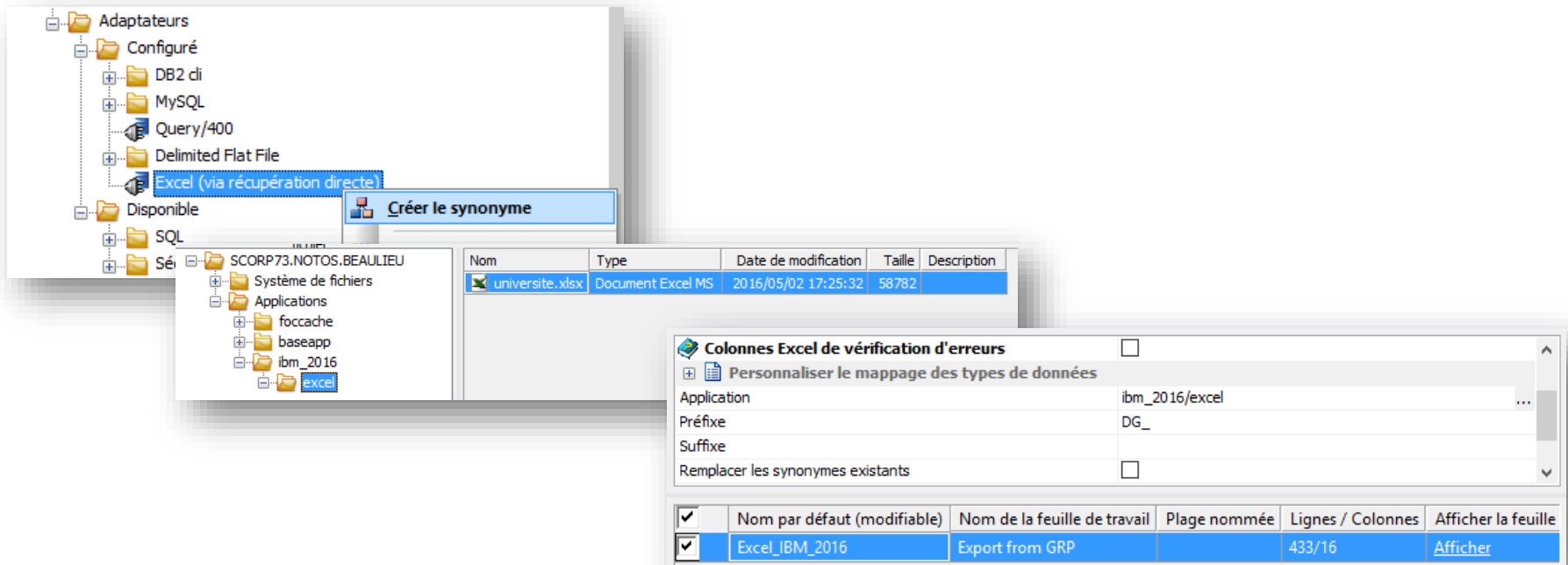
Conditions de la jointure :

	Colonne	Relation	Type	Valeur
1	T1.PRODUCTNUMBER	=	Colonne	T2.PRODUCTNUMBER

OK Annuler

Importation rapide (Quick ETL Copy)

- Importation de données en quelques clics
- Exemple à partir d'un fichier Excel importé dans une table de l'IBM i
- 1) Créer la métadonnées sur le fichier Excel



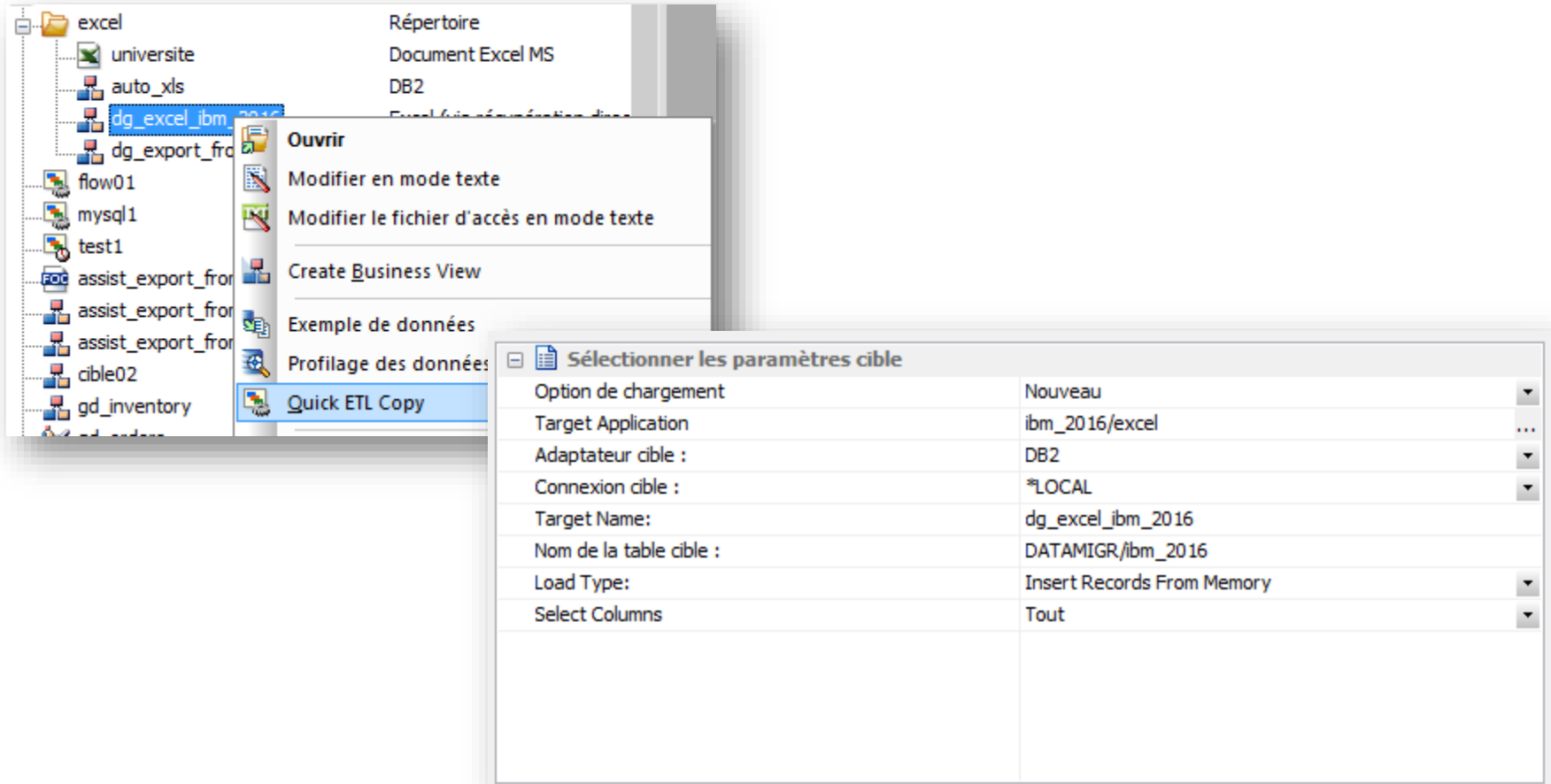
The screenshot illustrates the process of creating metadata for an Excel file in IBM i. It shows the 'Adaptateurs' tree with 'Excel (via récupération directe)' selected. A file browser displays 'universite.xlsx'. The 'Colonnes Excel de vérification d'erreurs' dialog box is open, showing a table of column mappings.

Nom	Type	Date de modification	Taille	Description
universite.xlsx	Document Excel MS	2016/05/02 17:25:32	58782	

	Nom par défaut (modifiable)	Nom de la feuille de travail	Plage nommée	Lignes / Colonnes	Afficher la feuille
<input checked="" type="checkbox"/>	Excel_IBM_2016	Export from GRP		433/16	Afficher

Importation rapide (Quick ETL Copy) (2)

- 2) Importer les données dans une table de l'IBM i



excel Répertoire
universite Document Excel MS
auto_xls DB2
dg_excel_ibm_2016
dg_export_fro

Ouvrir
Modifier en mode texte
Modifier le fichier d'accès en mode texte
Create Business View
Exemple de données
Profilage des données

flow01
mysql1
test1
assist_export_fro
assist_export_fro
assist_export_fro
cible02
gd_inventory

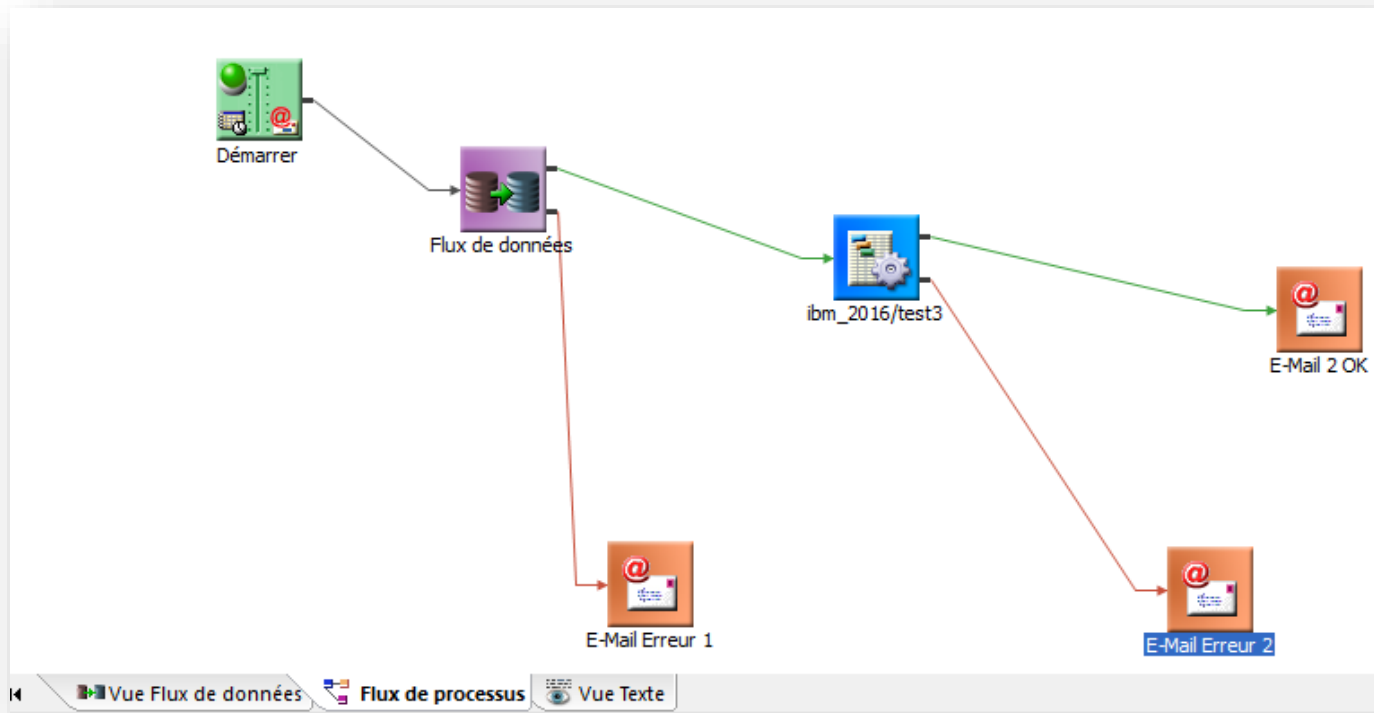
Quick ETL Copy

Sélectionner les paramètres cible

Option de chargement	Nouveau
Target Application	ibm_2016/excel
Adaptateur cible :	DB2
Connexion cible :	*LOCAL
Target Name:	dg_excel_ibm_2016
Nom de la table cible :	DATAMIGR/ibm_2016
Load Type:	Insert Records From Memory
Select Columns	Tout

Flux de processus

- Permet d'enchaîner des flux de données
- Définition de véritables « applications » complexes
- Gestion des erreurs et des évènements



Conclusions

- L'IBM i est une excellente solution pour héberger un Datawarehouse
 - Disponibilité
 - Volumétrie
 - Puissance de traitement
 - Administration
 - Sécurité
 - ...

- Avec Data Migrator for i on dispose d'un véritable ETL totalement intégré à l'IBM i
 - Accès à des bases de données variées
 - Traitement des fichiers plats (Microsoft Excel)

Merci pour votre écoute !

Des questions ?

Dominique GAYTE - dgayte@notos.fr
04 30 96 97 33
www.notos.fr