

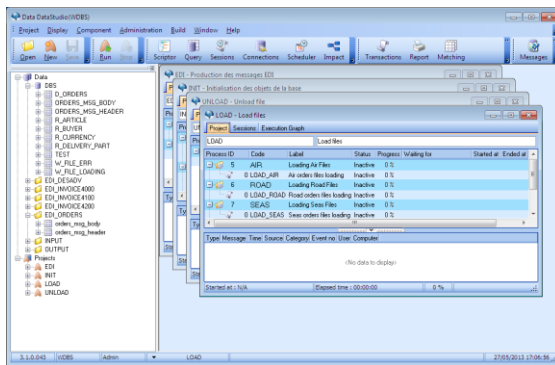
## Présentation du produit

### ■ Visualiser, manipuler et échanger les données

Le DataStudio est un ETL destiné aux architectures multi plates-formes. Facile d'accès, il s'appuie sur des instructions en SQL, le langage de requêtage de référence des informaticiens. De l'identification à la diffusion, le DataStudio aide à manipuler toutes les données de l'entreprise, dans tous les formats (du simple fichier Ascii, aux données des mainframes).

Le produit offre une représentation simple et complète des données et des traitements :

1. Référencement des données : Data
2. Mise en œuvre des traitements : Projects



### ■ Architecture Orientée Services « SOA »

Sur la base d'un moteur d'échange de données architecturé pour une approche orientée services « SOA », le DataStudio version 3.x devient une véritable plate-forme pour répondre de manière uniforme et organisée à l'ensemble des besoins d'intégration de données et de diffusion d'information de l'entreprise.

DataStudio dispose des fonctions suivantes :

- ✦ Fonctionnement en service,
- ✦ Exécution à distance des traitements,
- ✦ Serveur de licence,
- ✦ Service de Planification des traitements : le Scheduler

### ■ Liste des modules en standard

	Agent d'exécution		Programme de désinstallation
	Aide en ligne (en français et en anglais)		Proxy
	Atelier de Design		Scheduler
	Client Distant		SDK
	Client Local		Serveur de licences
	Service de supervision		Serveur de traitements

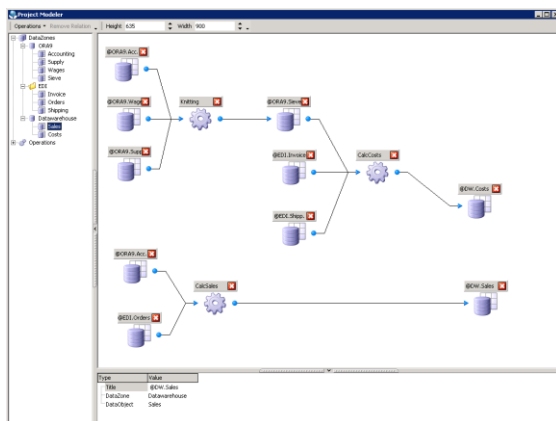
## Atelier de Design

### ■ L'interface de modélisation graphique des traitements

Le module Project Modeler facilite la mise en œuvre des projets et la communication entre les équipes.

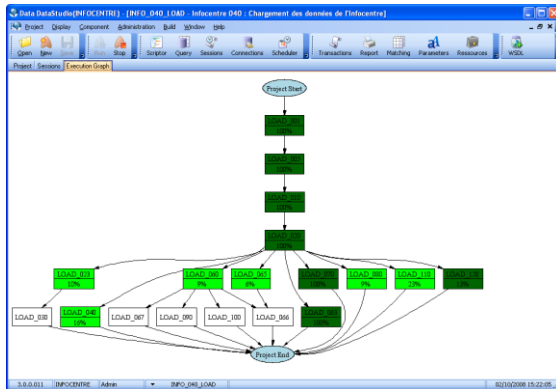
Les fonctions sont les suivantes :

1. Modélisation graphique,
2. Programmation des opérations en Sql,
3. Génération du projet.



Project Modeler

### ■ Graphe de l'arbre de traitement du Projet

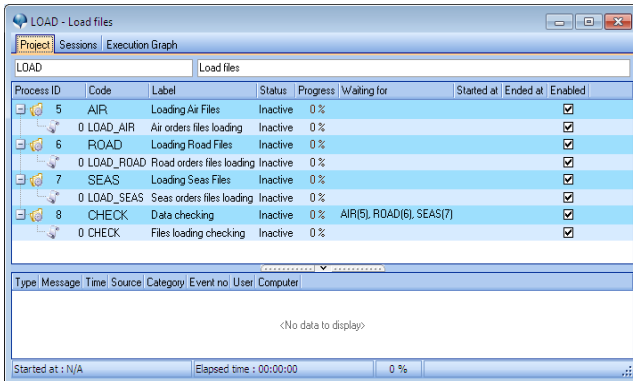


Graphe d'exécution

Ce graphe est généré à partir du paramétrage. Il permet :

- ✦ de comprendre la structure du traitement,
- ✦ de repérer d'éventuelles anomalies,
- ✦ de suivre visuellement l'exécution des traitements (le graphe est animé).

## ■ Projet DataStudio

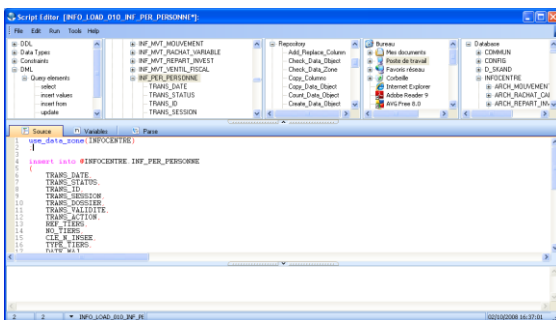


### Projet DataStudio

Le projet est l'unité de traitement de base du DataStudio.

L'écran d'exécution des projets permet de suivre en temps réel, l'exécution complète et détaillée par dossier et par job de chacun des projets : État, jauge de progression, heure de départ et de fin.

## ■ Module de création de traitements



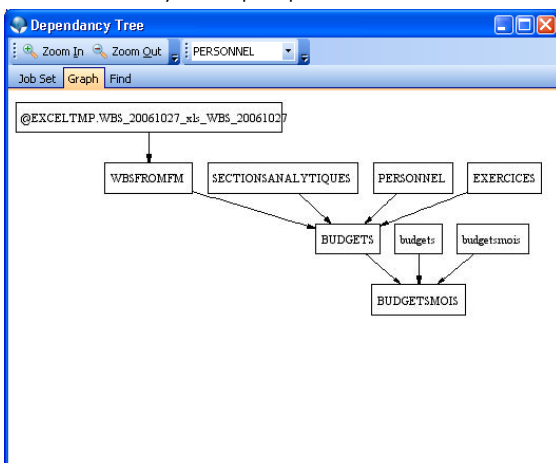
### Script Editor

L'éditeur de script a été renforcé des fonctions suivantes :

- ✦ Assistant de génération de scripts de création de base de données en drag&drop,
- ✦ Assistant de génération des mappings avec les messages web services,

## ■ Module d'analyse d'impact

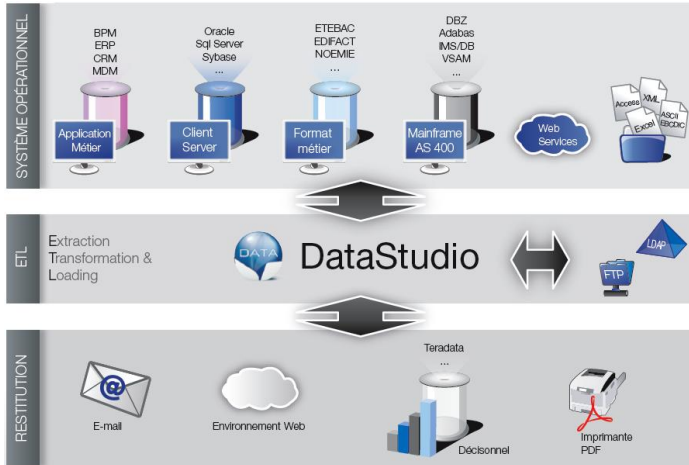
Le module d'analyse d'impact permet de suivre une donnée à travers le graphe des traitements.



### Dependency tree

## Le référencement des données

Dès sa mise en œuvre, le DataStudio élargit le champ d'action du service informatique. Le logiciel de DATA est à même de lister l'ensemble des données numérisées de l'entreprise. En effet, il s'interconnecte avec toutes les applications existantes (ou à venir) et il reconnaît tous les formats de données.



### ■ Description des données –métadonnées

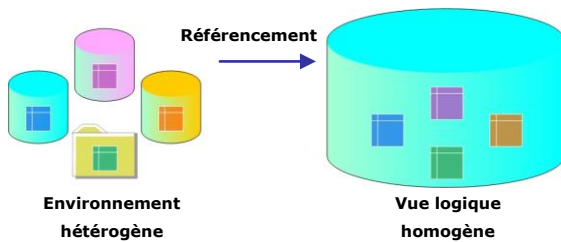
DataStudio répertorie toutes ces informations, indépendamment de leur format d'origine. Ces données sont ensuite présentées dans un espace de travail ou atelier de design, à l'aide d'une interface graphique homogène, lisible et logique. Ainsi, les données sont classées selon une nomenclature définie par l'entreprise, facilitant ainsi leur manipulation.

### ■ Définition de la donnée

Le DataStudio offre une représentation uniforme et complète de l'ensemble des données du système

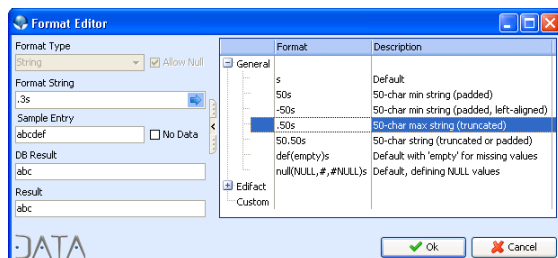
Les données sont représentées conformément au formalisme suivant :

- + Les sources de données sont nommées data zones (base de données, répertoire, url...),
- + Les données sont contenues dans les data objects (tables, fichiers, pages xml ou html...).



*Une fois référencées l'ensemble des données pourront être manipulées indépendamment de leur origine*

### ■ Ecran de conception de formats



Format editor

Cet écran permet :

- + de créer de nouveaux formats à partir d'un ensemble de formats modèles,

- ✦ d'enrichir une liste de formats personnalisés,
- ✦ de tester le bon fonctionnement du format par rapport à une chaîne à saisir,
- ✦ d'accéder aux formats spécifiques pré-chargés d'un métier (comme EDIFACT).

## ■ Ecran de visualisation et d'analyse

CONTRAT	SOUSCRIPTEUR	DATE_EFFET	DATE_RESILIATION	ETAT	AGE1	DATE_ECHEANCE	REF_EXTERNE
PRODUIT : 000							
00000173	11687	31/12/2002		V	(Custom.)	30/12/2042	xxxxxxxxxx
00000219	11775	06/02/2003		V	0	05/02/2043	xxxxxxxxxx
00000310	11951	22/05/2003		V	1	21/05/2043	xxxxxxxxxx
00000354	12027	30/07/2003		V	2	29/07/2043	xxxxxxxxxx
00000467	12245	04/12/2003		V	0	03/12/2019	xxxxxxxxxx
00000572	20193	20/05/2004		V	0	19/05/2044	xxxxxxxxxx
00000576	28203	28/05/2004		V	0	27/05/2012	xxxxxxxxxx
PRODUIT : 002							
PRODUIT : 009							
00900015	12821	17/07/2003		V	0	16/07/2013	xxxxxxxxxx
PRODUIT : 010							
01000019	12909	09/05/2003		V	2	08/05/2013	xxxxxxxxxx
01000035	12929	18/07/2003		V	2	17/07/2011	xxxxxxxxxx
01000036	12931	18/07/2003		V	2	17/07/2011	xxxxxxxxxx
PRODUIT : 011							
PRODUIT : 012							
PRODUIT : 015							
PRODUIT : 016							
016000141	128450	04/05/2004		V	0	03/05/2044	xxxxxxxxxx

### Data viewer

L'écran de consultation et analyse de données a été amélioré offrant les caractéristiques suivantes :

- ✦ Utilisation plus intuitive,
- ✦ Possibilité de regroupement sur une clé multiple,
- ✦ Export vers Excel du résultat,
- ✦ Gestion de l'impression du résultat,
- ✦ Module de filtrage personnalisable finement ;

## Manipulation des données

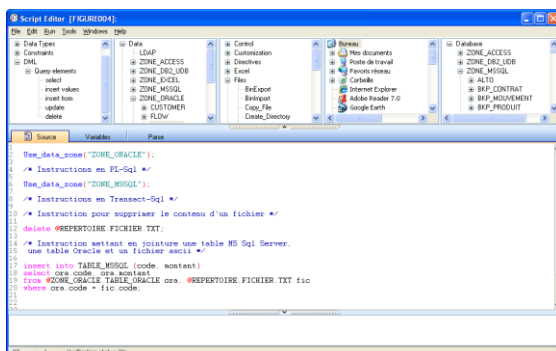
Une fois référencées dans le DataStudio, les données sont perçues comme appartenant à une base universelle. Il suffit alors au designer d'écrire des requêtes SQL pour les manipuler. En recourant à ce langage de programmation « naturel », l'informaticien est en mesure de réutiliser ses procédures ou ses modèles de requêtes existants.

## ■ Éditeur de script ergonomique

L'Éditeur de scripts « Script Editor », permet de programmer les instructions, et de les tester.

Le Script Editor dispose de toutes les fonctions de programmation et de tests :

- ✦ Aide à la saisie (wizard) des instructions et des ressources référencées,
- ✦ Analyseur de syntaxe en couleurs,
- ✦ Contrôle syntaxique global et pas à pas (avec jump et drill-down),
- ✦ Contrôle de cohérence des données,
- ✦ Console temps réel de suivi du traitement et des informations,
- ✦ Visualisation de l'interprétation du script et de la valeur des paramètres,
- ✦ Aide en ligne contextuelle et système de recherche par mot-clé,
- ✦ Recherche transversale dans tous les scripts par mot-clé avec accès par lien hypertexte,
- ✦ Import et export de scripts...



### ■ Langage de script avancé

Le DataStudio dispose d'un langage complémentaire du Sql de 3<sup>ème</sup> Génération, qui garanti la capacité de prise en charge de tous niveaux de complexité ou d'intégration avec les environnements en présences :

- ✦ Commandes systèmes,
- ✦ Manipulations des fichiers, Manipulations avancées Excel, XML,
- ✦ FTP, DDE, Pilotage d'ActiveX (OLE, COM),
- ✦ Chargement de DLL externes,
- ✦ Commandes procédurales (if, boucles, variables, curseurs, fonctions, paramètres, ...),
- ✦ Gestion des exceptions,
- ✦ Commandes de manipulation du référencement de données,
- ✦ Commandes de gestion d'événements sur les données,
- ✦ Envoi et réception des mails avec intégration des données,
- ✦ Import et export au format binaire (enregistrement de fichiers en base),
- ✦ Gestion de Transactions fédérant l'ensemble des données hétérogènes impliquées,
- ✦ Alimentation en mode Delta,
- ✦ Développement de modules de connaissance sur commande (Fonctions de pricing, Inversion de matrices, Tableaux croisés ...).

### ■ Langage de script universel

Le DataStudio propose un vocabulaire commun aux principaux moteurs de SGBD. Par exemple, le type de donnée « date\_\_ » du DataStudio sera interprété par « date » ou « datetime » à destination respectivement d'un moteur Oracle ou d'un moteur Sql Server ou Sybase.

En utilisant le vocabulaire générique associé à une syntaxe Sql Ansi 92 – norme sql commune à tous les SGBD – vous disposerez d'un script indépendant du type de SGBD utilisé (dans la limite du respect de cette norme).

Vous serez en mesure d'exécuter vos traitements ainsi constitués sur n'importe qu'elle base.

### ■ Contrôle de cohérence du référentiel de données

Le DataStudio dispose d'une interface de contrôle graphique de la cohérence des sources et de la description des données. La fonction d'analyse affiche distinctement l'ensemble des incohérences survenues (perte de connexion aux bases, inaccessibilité des répertoires, disparition ou simplement changement de la structure d'une table ou d'un fichier...).

Les fonctions de contrôle de cohérence peuvent être sollicitées automatiquement avant le lancement d'un traitement et l'interrompre avec alerte. Les contrôles sont aussi réalisés systématiquement au cours des traitements et gérés par le système de suivi des anomalies.

### ■ Nettoyage des données et suivi des anomalies

Au cours du chargement, le DataStudio détecte les données erronées et les aiguillent vers une zone de traitement des exceptions en y associant le code de l'erreur.

Dans cette zone, le designer définit sa stratégie de traitement de l'anomalie :

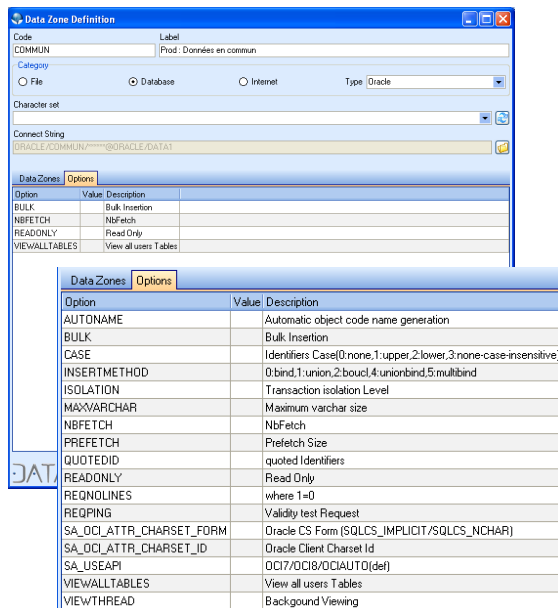
- ✦ Recyclage de ligne (valeurs par défaut, troncatures, Alerte...),
- ✦ Rejet avec ou sans stockage,
- ✦ Alerte temps réel,
- ✦ Interruption du traitement avec validation ou invalidation (rollback) de la session de chargement.

## Serveur de traitements

### ■ Performances améliorées

Les performances sont augmentées :

- ✦ Initialisation et terminaison plus rapide des projets,
- ✦ Transferts améliorés : parallélisation lecture/écriture,
- ✦ Gestion de la mémoire spécifique pour environnements parallèles sans goulots d'étranglement,
- ✦ Option Bulk étendue sur de nouveaux types de connexions,
- ✦ Option d'optimisation des connectivités aux Data Zones,



Options avancées des zones de données

### ■ Gestion de la qualité de données des fichiers

DataStudio gère finement les valeurs de substitution :

- ✦ Valeurs par défaut : en cas de valeurs manquantes,
- ✦ Valeur de remplacement en cas d'erreur de format,
- ✦ Valeurs personnalisables pour le NULL,

### ■ Gestion de la qualité des transferts

DataStudio capture et traite toutes les anomalies sur les données détectées :

- ✦ Fonctions de DataFlow avec gestion des exceptions par ligne avec des gestionnaires standard ou personnalisables,

### ■ Gestion de la qualité des données d'adresses

DataStudio a été enrichi de fonctions métier pour le traitement de la qualité des adresses :

- ✦ Fonction de qualité de données de normalisation et correction d'adresses de la poste française et internationale,

### ■ Fonctions de traitement standardisés

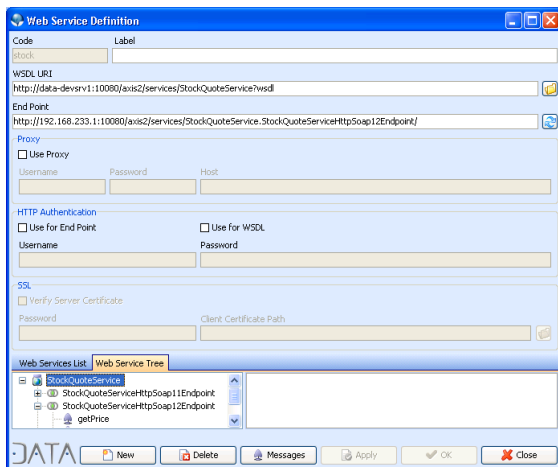
DataStudio permet la mise en œuvre de scripts génériques :

- ✦ Fonctions DataStudio utilisable directement dans les requêtes SQL afin de créer des instructions indépendantes des spécificités des langages SQL des fournisseurs de base de données.

## Web Services

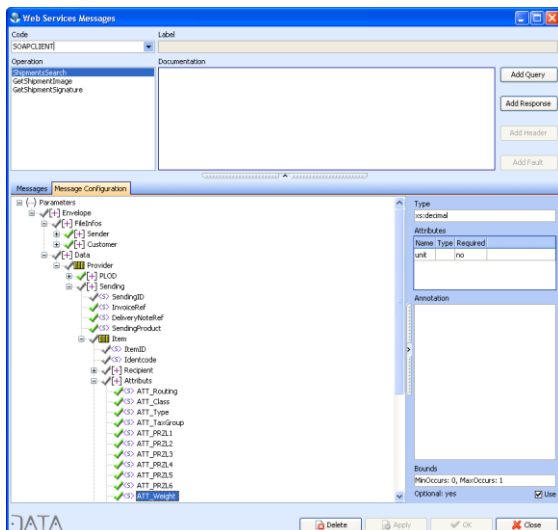
### ■ Ecrans de référencement des web services SOAP

DataStudio permet le référencement des Services par lien WSDL,



**Web service definition**

Ainsi que le référencement des messages SOAP par opération et sélection des éléments utilisés.



**Web services messages**

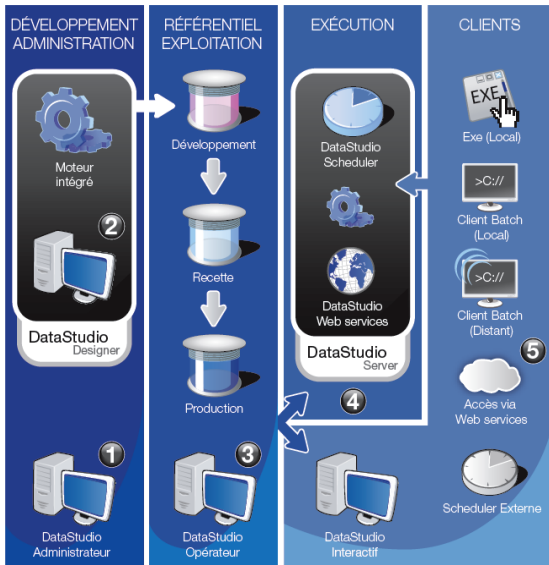
## Gestion des projets

### ■ Référentiels

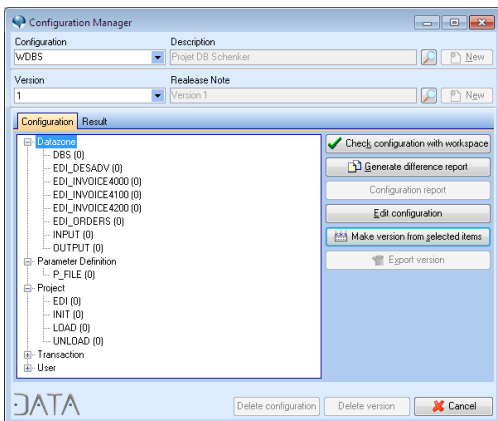
Les étapes principales de la mise en œuvre sont les suivantes :

1. Gestion des utilisateurs et des droits associés
2. Référencement de données & Design de projets – Versionning de la configuration,
3. Chargement des versions dans les référentiels de recette et de production – Paramétrage et activation du Scheduler,
4. Chargement du Workspace en mémoire,
5. Déclenchement et suivi des traitements





■ **Versioning des projets**



Le DataStudio dispose d'un module de gestion de la configuration. Les ressources livrables du Workspace sont intégrées à la configuration.

Les fonctions de gestion de la configuration sont notamment, les suivantes :

- ✦ Versioning de la configuration,
- ✦ Détection des écarts entre deux versions,
- ✦ Possibilité de réactiver une version précédente,
- ✦ Export et l'import contrôlé des configurations à l'aide d'un seul fichier XML,
- ✦ Détection et chargement automatique des nouvelles versions,
- ✦ Compte rendu automatique d'installation.

■ **Déploiement**

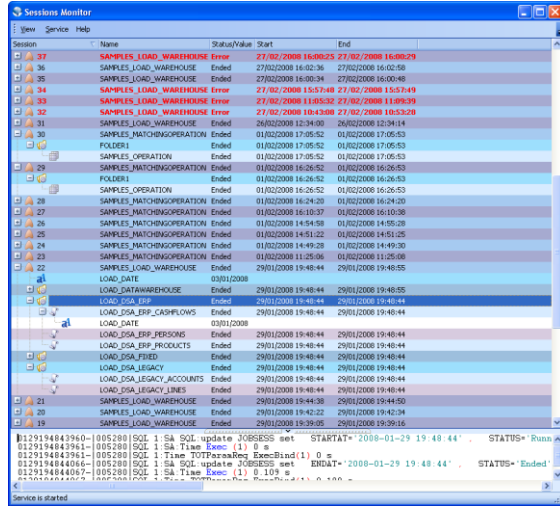
Le Workspace est enregistré dans une base de l'environnement librement choisie. Le moteur de traitement charge le référentiel sélectionné en mémoire et dispose de la dernière version des traitements.

Les traitements sont lancés par la technique fixée par l'utilisateur :

- ✦ Lancement en mode interactif à travers l'interface,
- ✦ Par un Client batch local (généralisé avec mot de passe crypté),
- ✦ A l'aide du Client batch distant,
- ✦ Via un accès web services (SOA),
- ✦ Lancement par le scheduler « maison » ou programmé dans le scheduler avancé DataStudio.

## Moniteur des logs

Le moniteur de suivi des traitements centralise toutes les informations enregistrées lors de leurs exécutions.



Moniteur de suivi des traitements

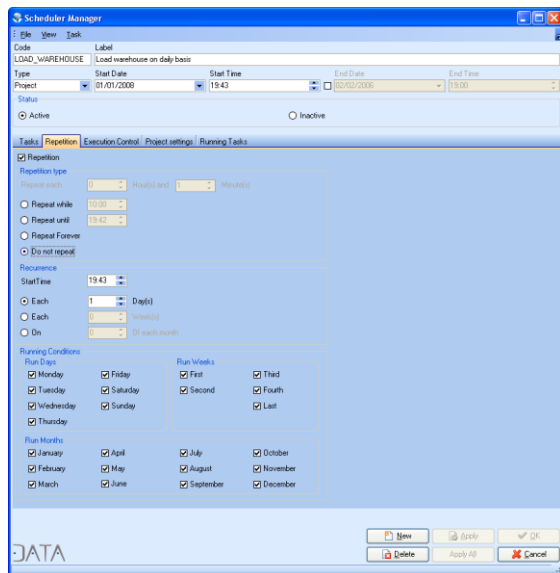
## Service Scheduler

Le Scheduler est un module est indépendant du DataStudio pour plus de robustesse

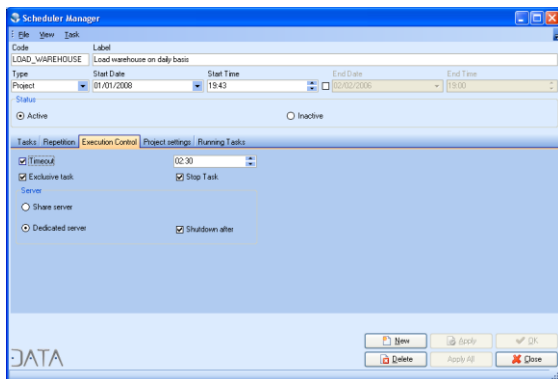
Toutes les finesses de traitements conditionnels peuvent être gérées soit au niveau des projets soit dans les règles du scheduler.

Le paramétrage du scheduler est enregistré dans un simple fichier permettant un déploiement aisé et une faible dépendance par rapport aux systèmes externes ou au serveur sources de dysfonctionnement. Le scheduler se caractérise par sa fiabilité.

Le paramétrage est mis en place à l'aide d'écrans intuitifs.



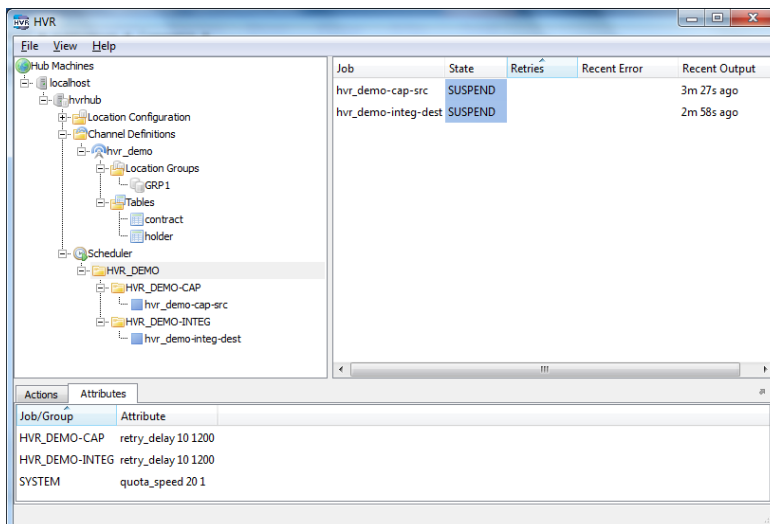
Scheduler manager – règles de répétitions



Scheduler manager – contrôle de l'exécution

## Change Data Capture

Le DataStudio intègre le module de réplication HVR pour capturer les changements au fil de l'eau dans les bases sources. Ces modifications sont alors répercutées en quasi-temps réel sur les bases cibles.



## SDK

### ■ Création de nouveaux formats de fichiers

- ✦ Environnement de développement d'extension aux formats fichier de DataStudio assisté,
- ✦ Performance car génération de binaires,
- ✦ Gestion de version des extensions,

## Serveur de licences

### ■ Gestion centralisée des droits

Le serveur de licence simplifie le déploiement du DataStudio à l'aide des fonctions suivantes :

- ✦ Gestion de l'affectation des droits personnalisable par le client,
- ✦ Nouveau type de licence flottante,
- ✦ Un seul lieu pour la gestion des licences,
- ✦ Repérage automatique du serveur de licence dans le réseau,
- ✦ Exécution en service

## Serveur de traitements

#### ■ API Web Services SOAP

- ✦ WSDL décrivant le service,
- ✦ Serveur http embarqué ne nécessitant aucun module supplémentaire,
- ✦ Sécurité par accès chiffrés par SSL.

#### ■ Service de supervision

- ✦ Garantie de la continuité de service en cas de défaillance
- ✦ Suivi de la bonne santé du serveur,

## Une approche ELT pour les meilleures performances

DataStudio fonctionne en standard en mode ELT par génération d'instructions exécutées par les moteurs des bases de données mises en jeu. A l'inverse la plupart des autres produits fonctionnent en mode ETL imposant l'utilisation d'un moteur propriétaire par lequel devront transiter les données à la moindre manipulation.

## Modules supplémentaires

#### ■ Module de reporting intégré

Le DataStudio intègre en standard, les fonctions d'appels des reports Crystal Report, Birt et BDoc.

Le DataStudio prend en charge la préparation des données et délègue à la mise en forme et la présentation du report à l'outil de reporting,

Cette approche présente plusieurs avantages :

- ✦ Facilité de réalisation,
- ✦ Performances très largement améliorées (temps de réponse),
- ✦ Bénéfice de la qualité de design et de présentation de l'outil de reporting,
- ✦ Sortie paramétrable dynamiquement de l'état vers l'écran, l'imprimante, au format pdf ou html,
- ✦ Déploiement immédiat et automatique,
- ✦ Pilotage du report avec passage de paramètres utilisateur et systèmes (chaîne de connexion notamment).

#### ■ Rapprochement de données

Le rapprochement de données – ou matching – consiste à comparer de sources de données et à établir le compte rendu des différences,

Le DataStudio s'appuie sur ses capacités d'accès aux données pour comparer automatiquement le contenu de deux tables, une table et un fichier ou deux fichiers quelle que soit leur origine ou leur localisation,

Le scénario de matching est basé sur les relations établies entre les deux sources, déterminant les champs qui servent de clés d'identification des lignes équivalentes, ainsi que la liste des colonnes à comparer.

Le traitement produit un compte-rendu détaillé du rapprochement avec :

- ✦ Les lignes matchées,
- ✦ Les lignes isolées (non trouvées dans l'autre source),
- ✦ Les lignes en écart et les valeurs respectives.

Le compte rendu est enregistré en base, disponible en consultation et éditable. Il est possible de définir un job de traitement des anomalies ou d'alertes.

## Connectivité

Configuration requise et plates-formes prises en charge.

#### ■ ETL Designer :

- ✦ Windows.
- ✦ Linux en mode dégradé Wine

#### ■ ETL Moteur de traitements :

- ✦ Windows.
- ✦ Linux
- ✦ Sun Solaris
- ✦ IBM AIX
- ✦ HP-UX 11 (en cours de certification)

#### ■ ETL Référentiel :

- ✦ Mysql
- ✦ Interbase/Firebird,

- ✦ PostgreSQL,
- ✦ Informix,
- ✦ MS SQL Server,
- ✦ Oracle,
- ✦ Sybase SQL Anywhere,
- ✦ Sybase SQL Server,
- ✦ Sybase Adaptive Server,
- ✦ IBM UDB.

#### ■ **Connectivité bidirectionnelle avec les formats de fichiers plats :**

- ✦ Fichiers en format texte délimité/fixe,
- ✦ Fichiers plats AS-400,
- ✦ Fichiers XML,
- ✦ Fichiers plats standards (Cobol),
- ✦ Fichiers plats exotiques (module de personnalisation tout format)

#### ■ **Connectivité avec les formats de fichier Microsoft :**

- ✦ Excel (bidirectionnelle)
- ✦ Word (rtf)
- ✦ Power point

#### ■ **Connectivité bidirectionnelle avec les SGBDR :**

- ✦ Microsoft SQL Server (native/SQL, ODBC)
- ✦ Microsoft Access (ODBC)
- ✦ Oracle (native, ODBC)
- ✦ Informix (native/SQL, ODBC)
- ✦ ODBC générique
- ✦ JDBC
- ✦ DBI (perl)
- ✦ API JSON (No Sql)
- ✦ Sybase (native, ODBC)
- ✦ Sybase IQ (native/SQL, ODBC)
- ✦ Sybase SQL Anywhere (native/SQL, ODBC)
- ✦ Teradata V2R3, V2R4 (native/SQL, ODBC)
- ✦ IBM UDB5, UDB6, UDB7 (native/SQL, ODBC)
- ✦ Mysql
- ✦ Interbase/Firebird
- ✦ PostgreSQL

#### ■ **Connectivité mainframe (Connecteur Attunity, ODBC) :**

- ✦ IBM DB2 AS400
- ✦ IBM DB2 MVS — OS/390
- ✦ IBM IMS
- ✦ IDMS
- ✦ Adabas C Versions 6.2 et 7.1
- ✦ VSAM KSDS et ESDS
- ✦ MVS Tape
- ✦ Fichiers AS/400
- ✦ C-ISAM

#### ■ **Connectivité avec les annuaires LDAP :**

- ✦ Protocole version 2, 3

#### ■ **Nouvelle connectivité à la base SAS**

DataStudio accède à la base SAS :

- ✦ DataStudio intègre le connecteur ODBC d'accès à la base SAS.
- ✦ DataStudio dispose aussi mode écriture Bulk offrant les meilleures performances en écriture dans la base SAS.