

# Présentation de l'ETL DataStudio

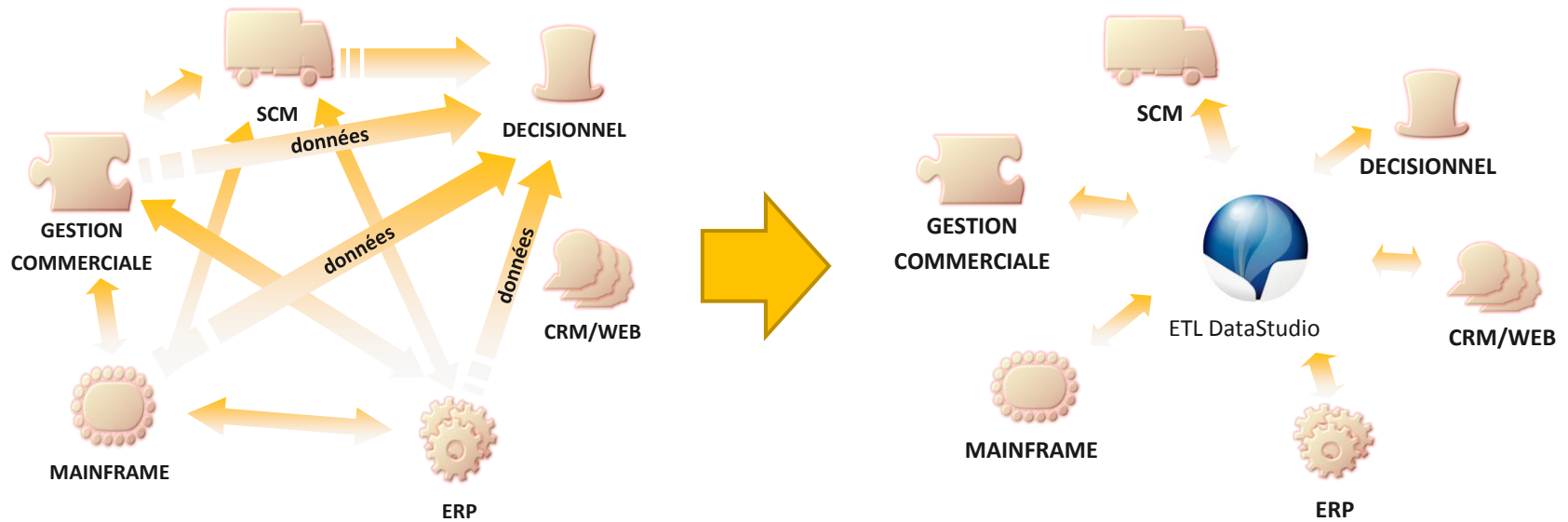
Solution d'intégration des données et  
de diffusion de l'information

Avril 2017

# Sommaire

- Simplifier les échanges de données
- La différence
- Référencement des données
- Manipulation des données
- Exemple de mise en œuvre
- Les bénéfices
- Les web services
- Change Data Capture
- CDC exemple
- DataStudio est un ELT
- Richesse fonctionnelle
- Performances

# Simplifier les échanges de données



- ETL : Extract, Transform & Load
- Échanges de données en environnement hétérogène
- Une alternative efficace aux développements spécifiques

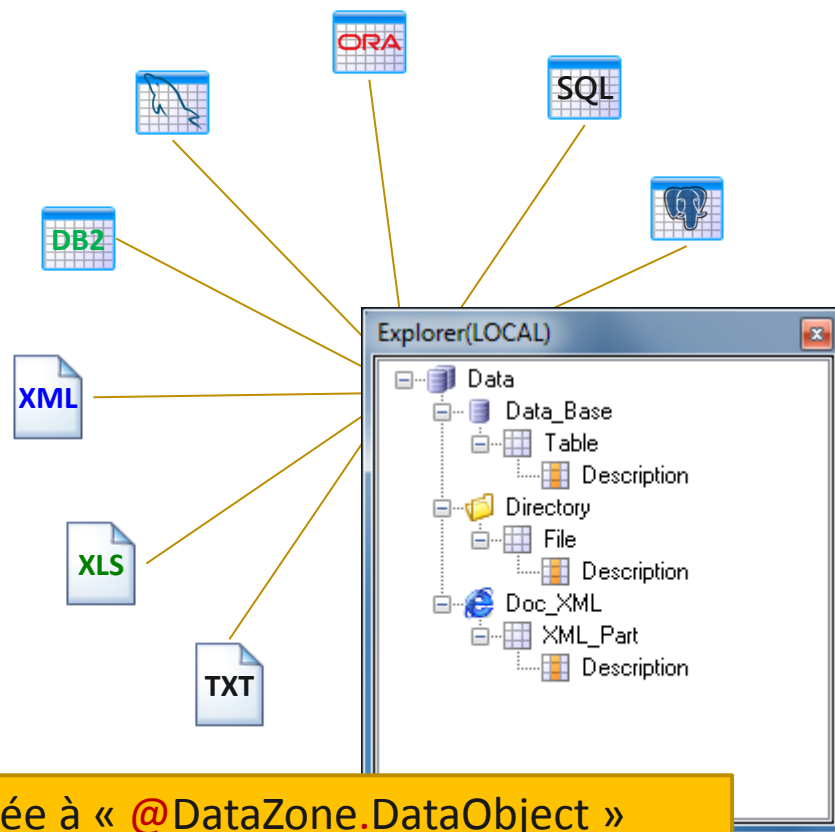
# La différence

- Pas de serveur dédié
- Plate-forme « tout-intégré » pour un travail en équipe efficace et sécurisé
- Prise en main rapide : tout en SQL

# Référencement des données

Formalisme simplifié :

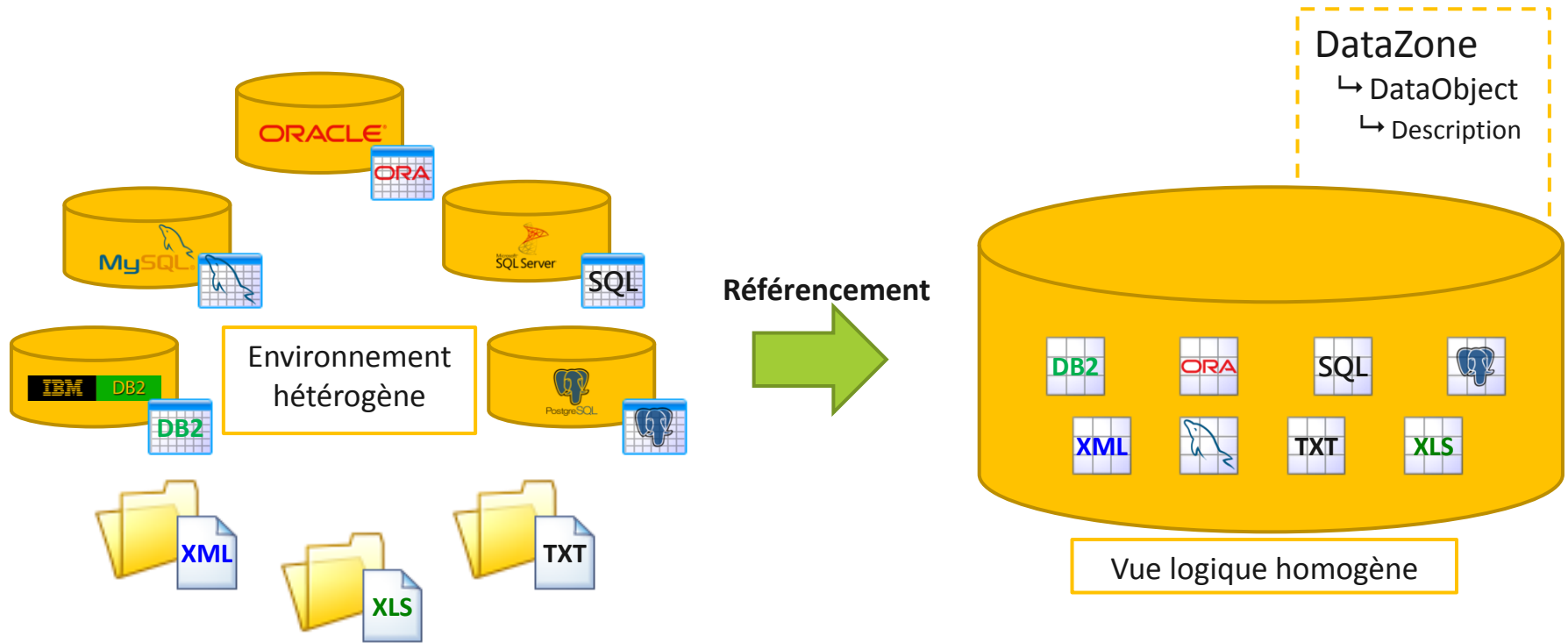
- Les sources de données sont nommées **DATA ZONES** (Base de données, Répertoire, URL...)
- Les données sont contenues dans les **DATA OBJECTS** (Tables, Fichiers, Pages xml ou html...)



Syntaxe **SQL** appliquée à « @DataZone.DataObject »

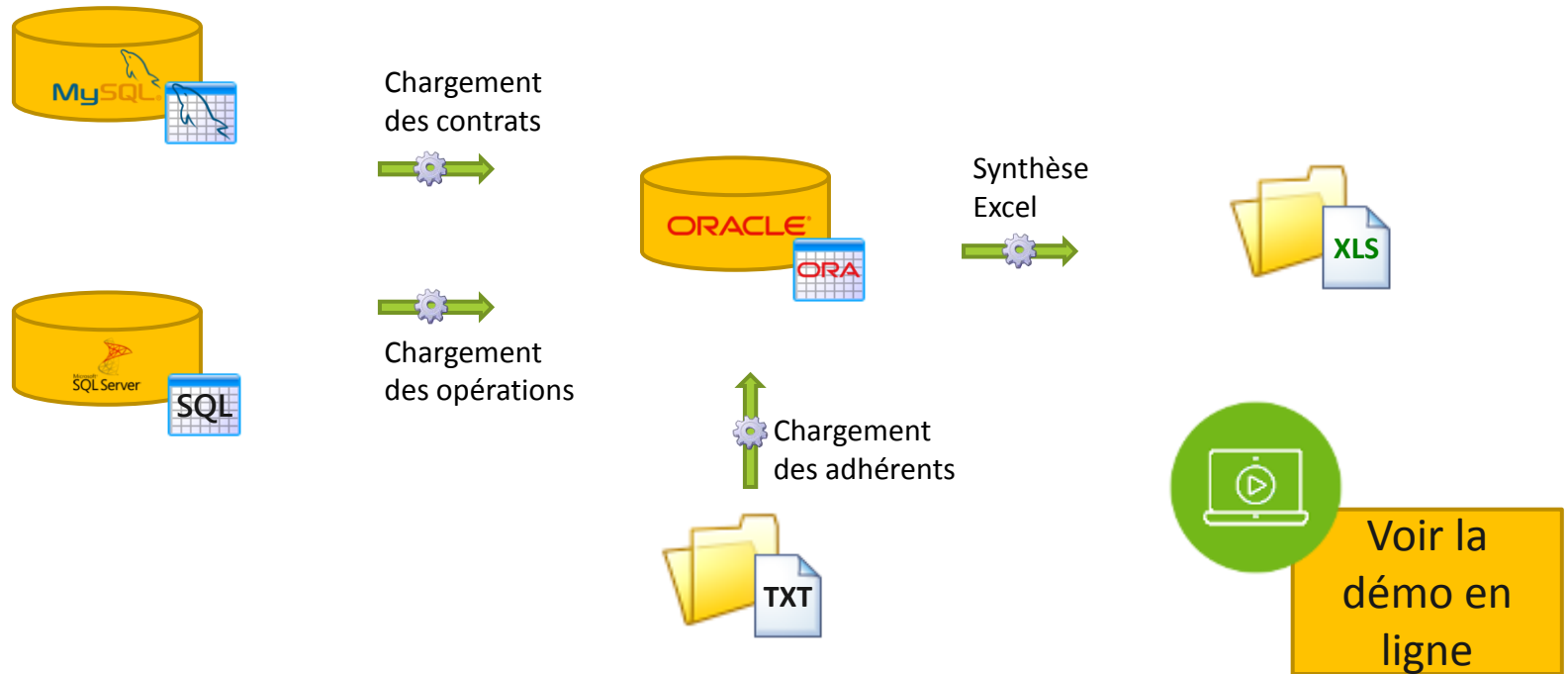
Représentation uniforme et complète de l'ensemble des données du système

# Manipulation des données



Quelle que soit l'origine des données, elles seront manipulées en **SQL** comme si elles étaient enregistrées dans les tables d'un même SGBDR

# Exemple de mise en œuvre



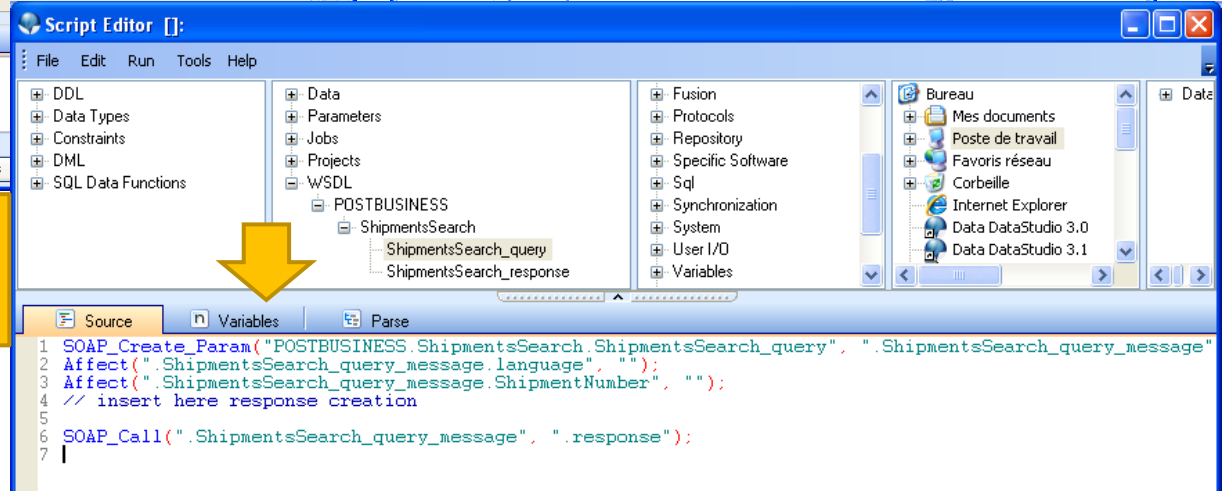
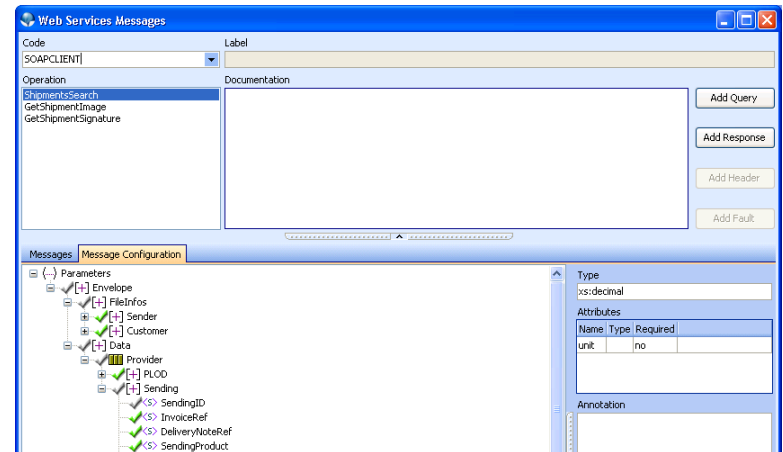
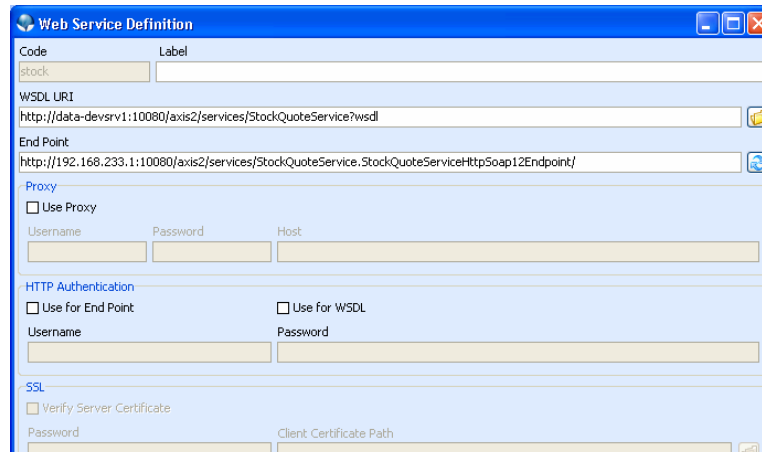
Approche originale de l'ETL DataStudio pour le référencement et la manipulation des données.

# Les bénéfices

- **Pas de serveur dédié**
  - Économique,
  - Performant.
- **Plate-forme « tout-intégré » pour un travail en équipe efficace et sécurisé**
  - Développement, Validation, Exploitation,
  - Suivi complet, traçabilité, accès sécurisés par rôles.
- **Prise en main rapide**
  - Tout en SQL,
  - Manipulation des technologies en place.
- **Retour sur investissement dès le premier projet**
  - Reprise de l'existant,
  - Gain de productivité,
  - Tarifs attractifs

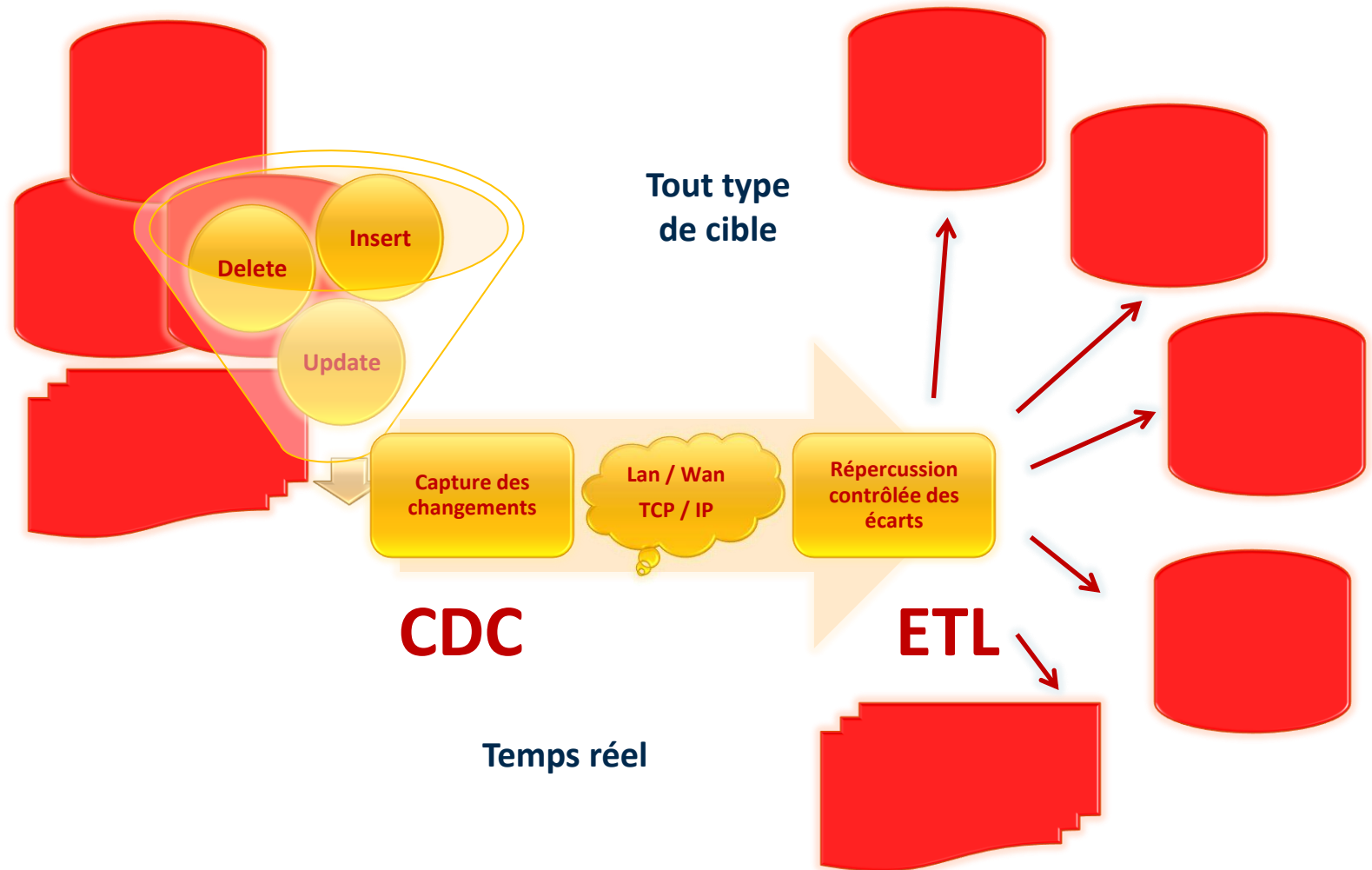


# Les web services pour une interopérabilité totale

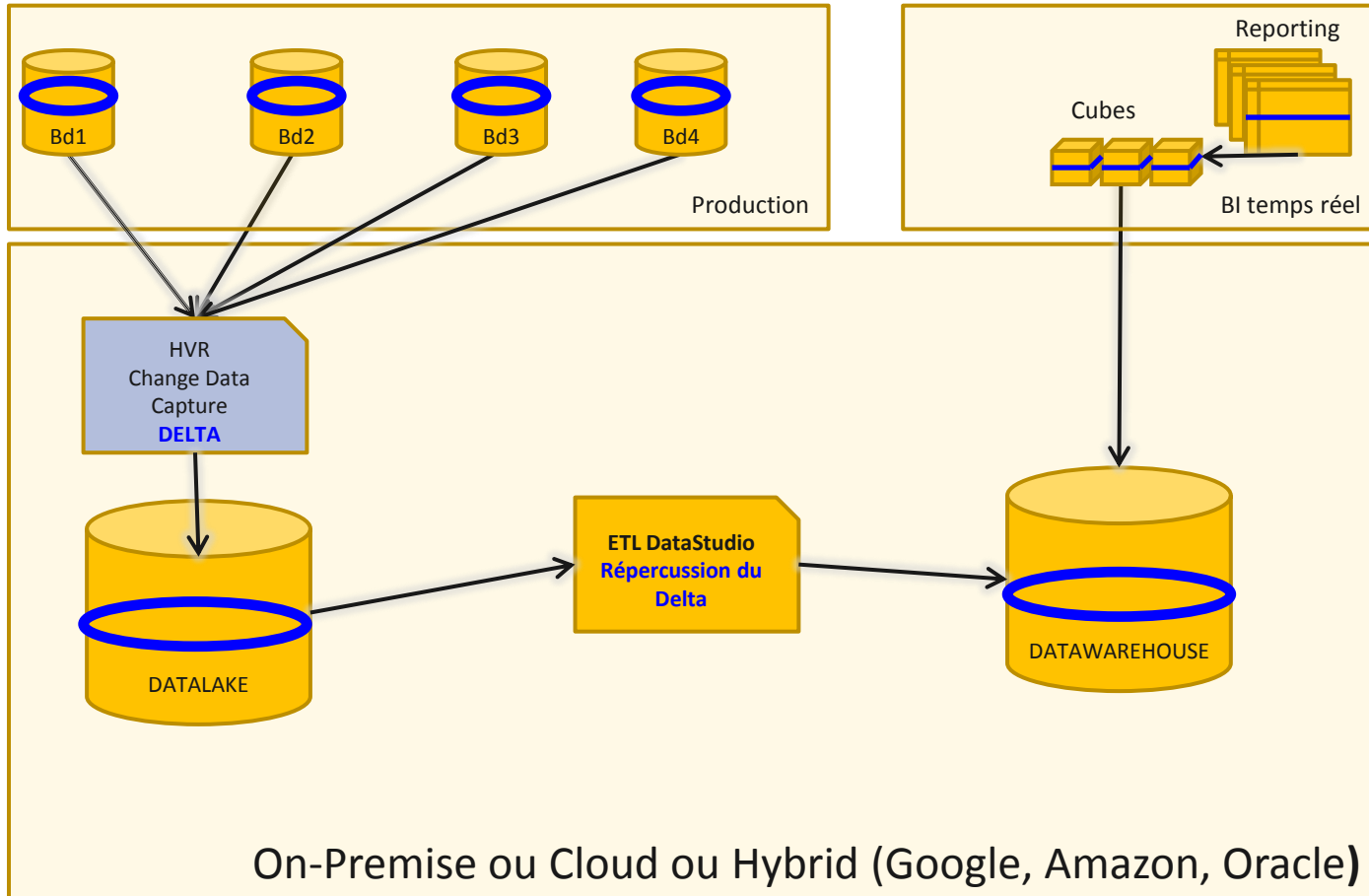


Mise en oeuvre par simple drag & drop

# Change Data Capture



# CDC exemple



# DataStudio est un ELT

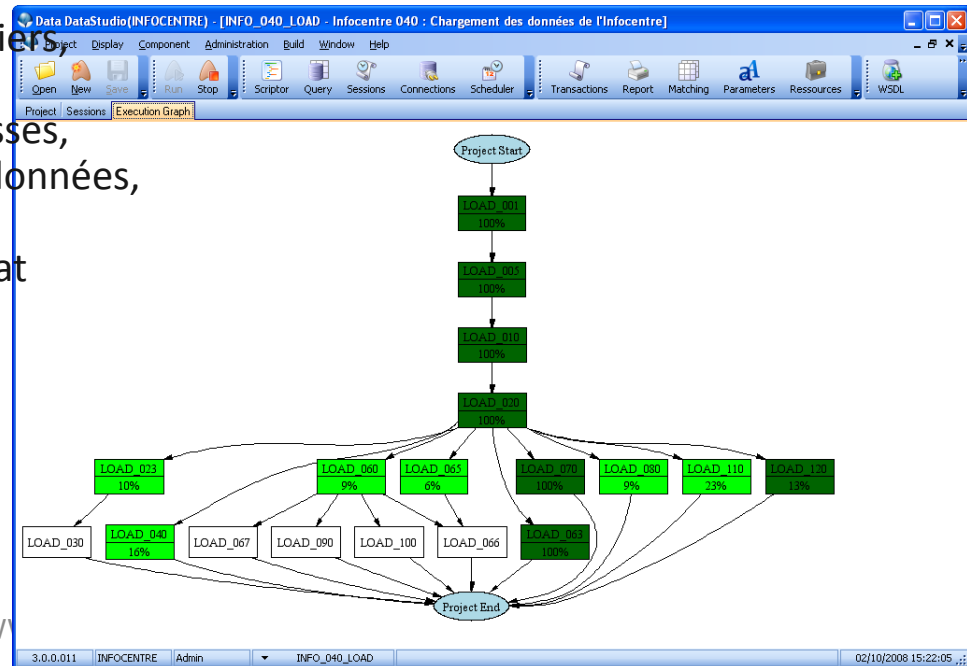
Montée en mémoire de toutes les données et manipulation dans la boucle



E.L.« T » (DataStudio)	E.« T ».L
<ul style="list-style-type: none"> <li>⊕ Coûts faibles,</li> <li>⊕ Performances optimales,</li> <li>⊕ Bénéfices des propriétés sécurisantes du SGBDR,</li> <li>⊕ Scalabilité et maintenance déjà prise en charge au niveau de la base,</li> <li>⊕ Rationalisation des ressources logicielles, matérielles et humaines du service informatique.</li> </ul>	<ul style="list-style-type: none"> <li>⊖ Coûts d'acquisition, de maintenance et d'exploitation élevés,</li> <li>⊖ Moteur propriétaire,</li> <li>⊖ Faibles performances,</li> <li>⊖ Risques de perte du fil de la transformation par le passage par des environnements hétérogènes de transformations.</li> </ul>

# Richesse fonctionnelle

- La plate-forme DataStudio dispose de nombreux modules pour répondre à tous les besoins de l'intégration des données :
  - Gestion des configuration et versioning,
  - Interface de comparaison entre deux configurations,
  - Assistant d'installation et de contextualisation des versions de configuration,
  - Graphe de suivi des traitements,
  - Contrôle de cohérence du référentiel de données,
  - Nettoyage des données et suivi des anomalies,
  - Écran de visualisation et d'analyse des données,
  - Requêteur multi-bases,
  - Module d'analyse d'impact,
  - Gestion de la qualité de données des fichiers,
  - Gestion de la qualité des transferts,
  - Gestion de la qualité des données d'adresses,
  - Module de comparaison des sources de données,
  - Module de reporting intégré
  - Diffusion de l'information sous tout format
  - Module de Change Data Capture,
  - ...



# Performances

- **Haut niveau de performance atteint dans le cadre des échanges de données :**
  - Performance basée sur l'architecture EL-T, la parallélisation et les technologies de chargements massifs de données
  - Moteur optimisé en code natif : aucun gaspillage de ressources
  - Connectivité standard diversifiée et native aux données (tout SGBD, tout format de fichier)

The screenshot shows the 'Data Zone Definition' dialog box. The 'Code' field is 'COMMUN' and the 'Label' is 'Prod : Données en commun'. The 'Category' is 'Database' and the 'Type' is 'Oracle'. The 'Connect String' is 'ORACLE/COMMUN/\*\*\*\*\*@ORACLE/DATA1'. Below this, there are two tables showing 'Data Zones' and 'Options'.

Option	Value	Description
BULK		Bulk Insertion
NBFETCH		NbFetch
READONLY		Read Only
VIEWALLTABLES		View all users Tables

Option	Value	Description
AUTONAME		Automatic object code name generation
BULK		Bulk Insertion
CASE		Identifiers Case(0:none,1:upper,2:lower,3:none-case-insensitive)
INSERTMETHOD		0:bind,1:union,2:boucl,4:unionbind,5:multibind
ISOLATION		Transaction isolation Level
MAXVARCHAR		Maximum varchar size
NBFETCH		NbFetch
PREFETCH		Prefetch Size
QUOTEDID		quoted Identifiers
READONLY		Read Only
REQNOLINES		where 1=0
REQPING		Validity test Request
SA_OCI_ATTR_CHARSET_FORM		Oracle CS Form (SQLCS_IMPLICIT/SQLCS_NCHAR)
SA_OCI_ATTR_CHARSET_ID		Oracle Client Charset Id
SA_USEAPI		OCI7/OCI8/OCIAUTO(def)
VIEWALLTABLES		View all users Tables
VIEWTHREAD		Background Viewing

Merci de votre attention

Retrouvez-nous sur [www.data.fr](http://www.data.fr)