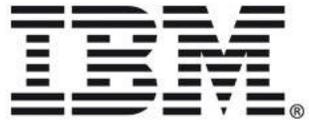


Tendances Logicielles
Hiver 2006





*Urbaniser les flux
d'information autour
de votre mainframe
Gautier Rose*



Agenda

- **Contexte et Enjeux**
- Modèles d'intégration
- Cas Client
- Synthèse

Les données mainframe dans le paysage informatique d'aujourd'hui

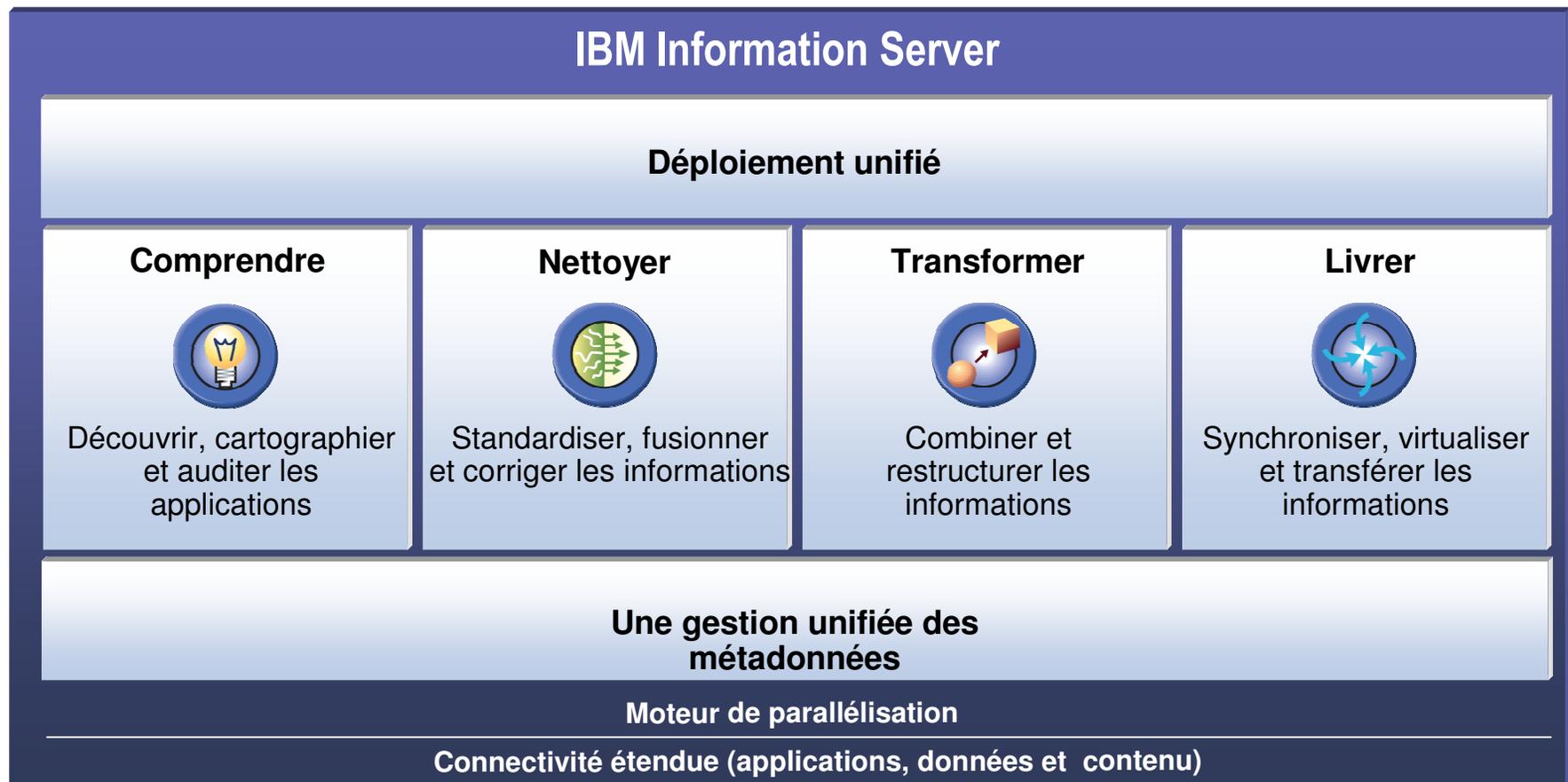
Au coeur du système métier

- **Données critiques** à l'entreprise (Clients, Produits, ...)
 - ▶ 60 % du volume des données, En croissance de 20 % par an.
- **Complexes** (VSAM, IMS, ADABAS, CA-IDMS, CA-DATACOM), **Volumineuses**
- De **qualité variable** en raison de **règles de gestion** qui ont évoluées dans le temps
- Dont la signification ou l'organisation est parfois **mal connue** ou **oubliée** (Fusion, turn-over, ...)
- Devant être **accessibles** au plus grand nombre **sous un délai minimum**

Objet d'initiatives majeures

- Des Projets réglementaires (BALE II, SOX, HIPAA, ...) ou Métiers (Vison Client 360°, Optimisation des stocks, ...) imposent une intégration **plus grande et plus fiable**.
- Avec des **sources et des cibles toujours plus volumineuses situées sur le mainframe**
 - ▶ Datawarehouse DB2 z/OS ou Teradata Channel-attached
- Et imposant une **réduction des coûts de développement et de maintenance**

IBM Information Server : *Pour un accès à une information fiable*

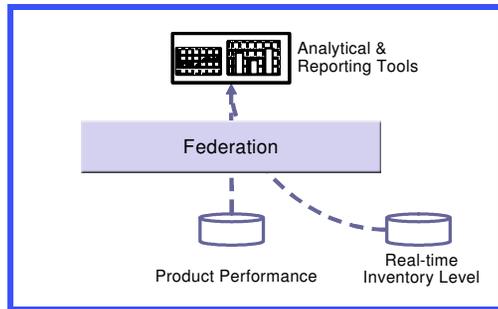


Agenda

- Contexte et Enjeux
- **Modèles d'intégration**
- Cas Client
- Synthèse

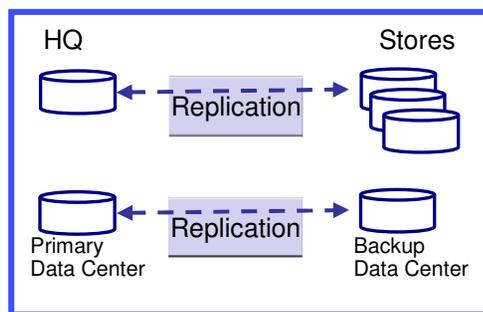
Plusieurs modèles d'intégration pour des besoins différents

Fédération



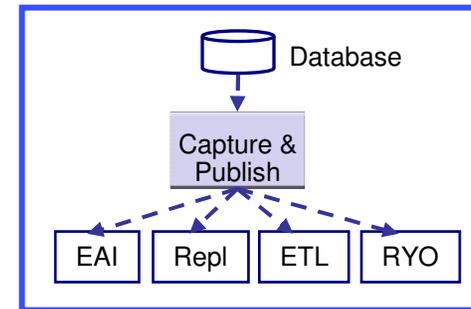
Accès temps réel sans programmation mainframe

Réplication



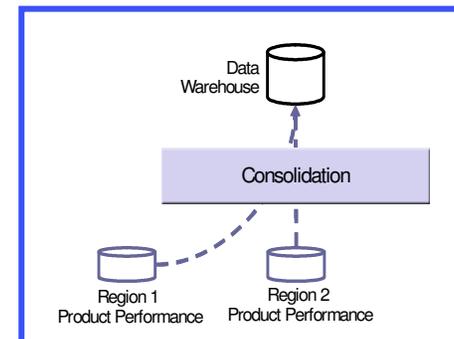
Assurer un disponibilité continue aux applications mainframe critiques

Publication



Mise à disposition immédiate des données opérationnelles aux applications métiers

Transformation

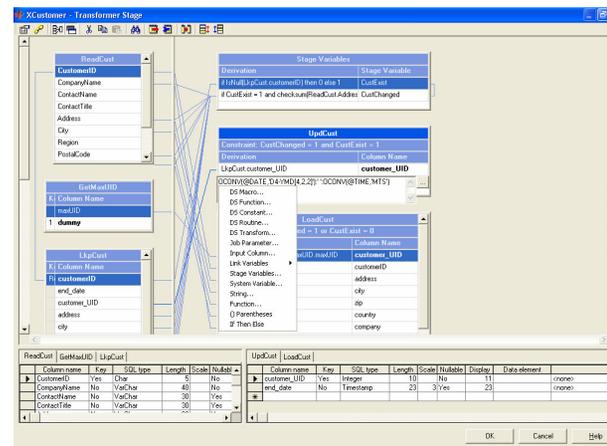
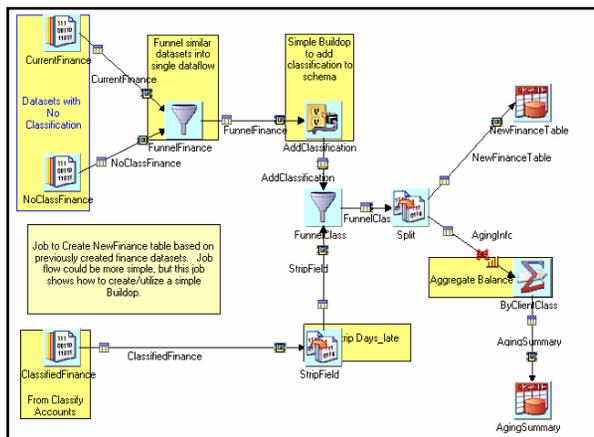


Réduire les coûts de développement des flux de données complexes et volumineux



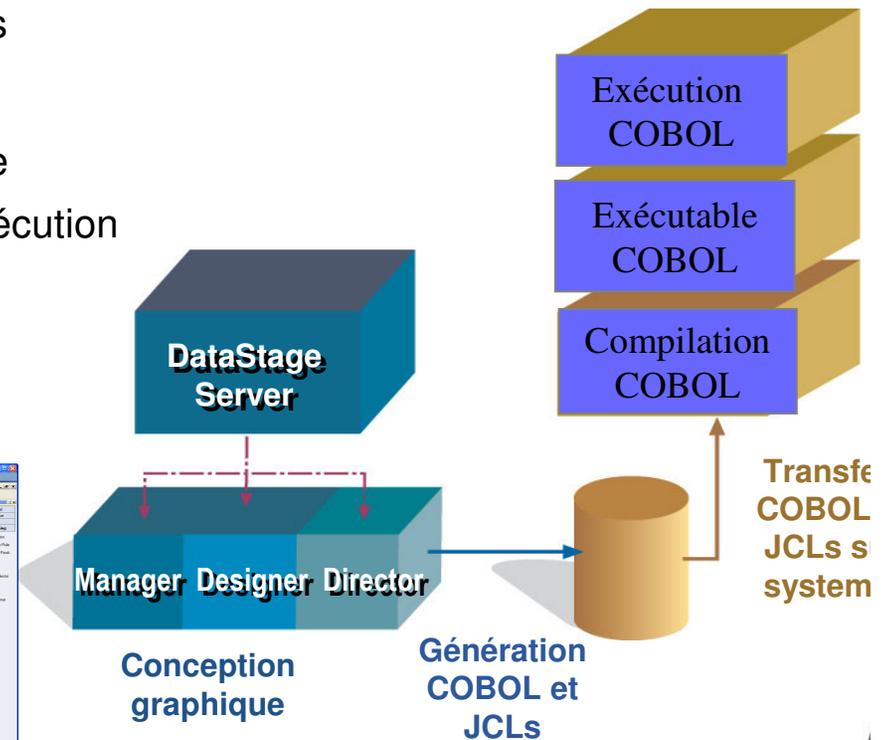
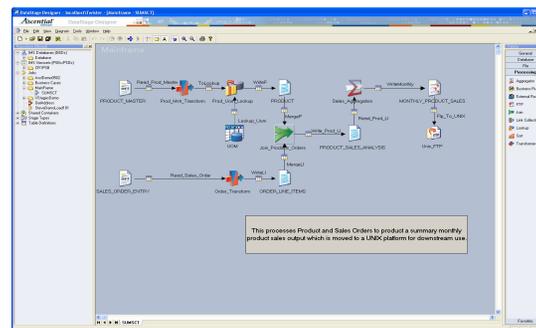
Transformation - DataStage

- **Modélisation graphique des flux de données**
 - ▶ Glisser-déplacer à partir d'une palette de composants génériques
 - ▶ Plus de 50 composants génériques et personnalisables :
 - Sources / Cibles : DB2, Oracle, SQL Server, XML, SAP, PeopleSoft, ...
 - Transformation : Fusion, Pivot, Agregation, Tri, SCD, ...
 - Réutilisation existant : procédures stockées, programmes Cobol, Java, SAS
 - ▶ Plus de 300 fonctions standards de transformation : manipulation de chaîne de caractères, calculs numériques et de dates, conversion de formats, ...
- **Compréhension simple, Rapidité de mise en œuvre**



Transformation – Mode Générateur COBOL

- Un programme COBOL est généré ainsi que des scripts JCL.
- Le programme COBOL est transféré sur system Z, puis compilé et exécuté à l'aide des scripts JCLs générés.
- Bénéfices :
 - ▶ Facilité d'intégration, Exécution native
 - ▶ Productivité accrue et Coûts de maintenance réduits
- Inconvénients :
 - ▶ Pas d'exécution parallèle : montée en charge limitée
 - ▶ TSO est toujours nécessaire à la compilation et l'exécution
 - ▶ Pas de portabilité des développements

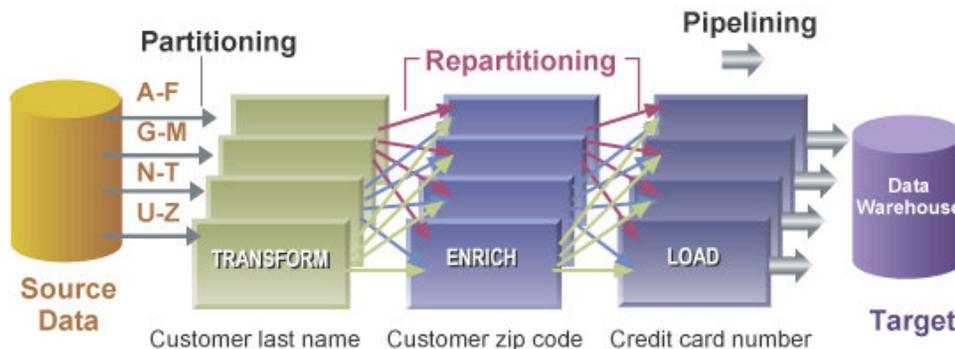


Transformation – Mode Moteur de transformation

L'utilisateur définit la logique fonctionnelle indépendamment de l'architecture technique



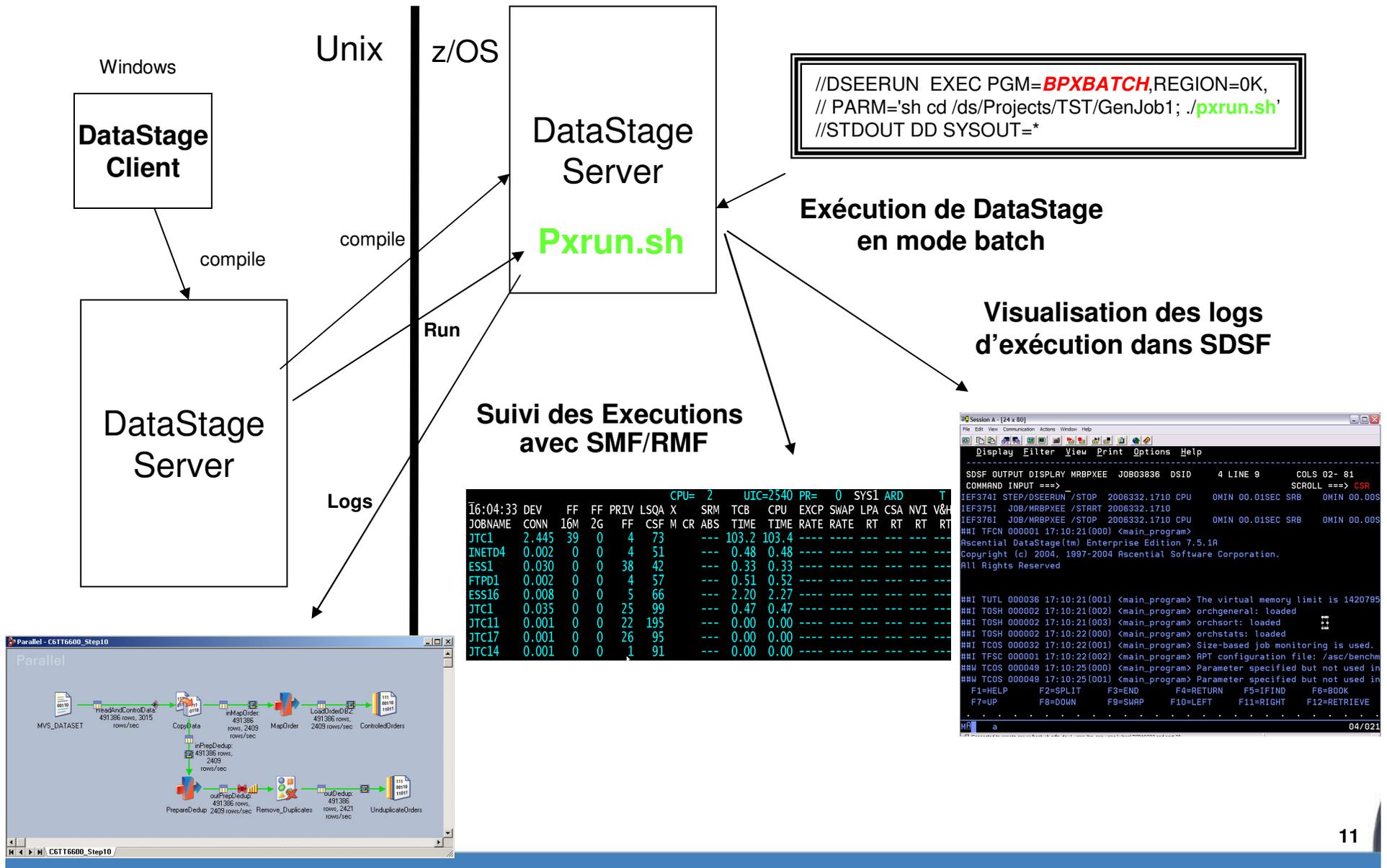
... A l'exécution, la logique fonctionnelle s'exécute en parallèle de façon transparente (1 CPU, 2 CPU, N CPU)



Un fichier de configuration décrit la topologie d'exécution.

Assure la montée en charge linéaire des volumes sans modification du Design

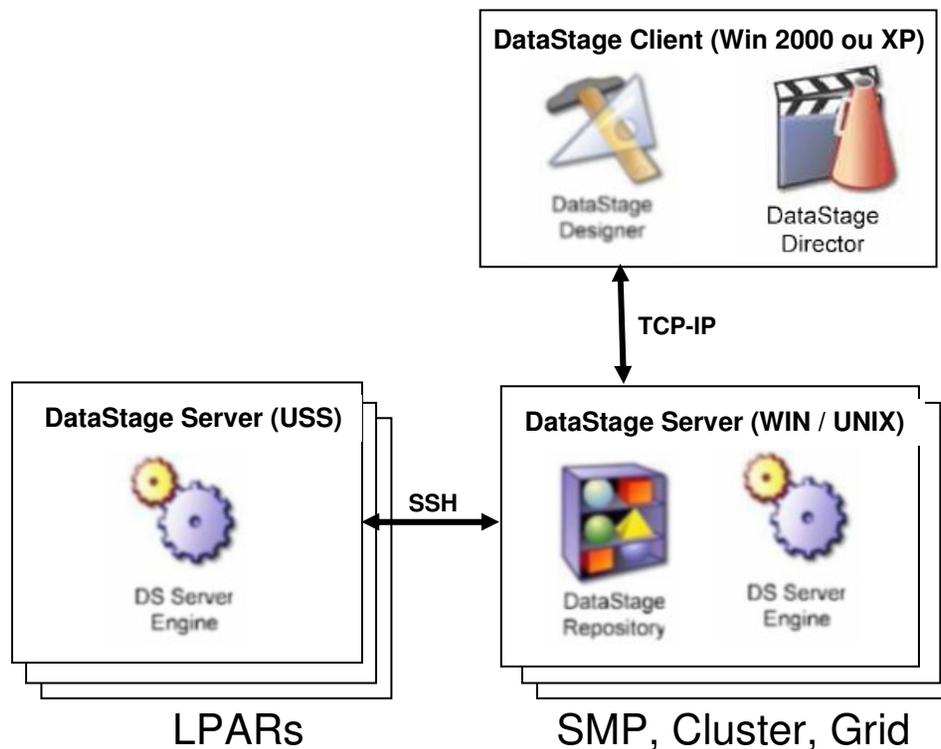
Transformation – Mode Moteur de transformation



Transformation – Moteur de transformation

- **Une même expérience quelque soit l'environnement**

- ▶ Un même environnement de développement graphique (Windows)
- ▶ Une même technologie de transformation (WINDOWS, AIX, SOLARIS, HP-UX, USS)



- **Bénéfices :**

- ▶ Haute Productivité
- ▶ Montée en charge sans limite et sans impact sur les développements
- ▶ Support natif z/OS
- ▶ Portabilité Unix / Windows / USS

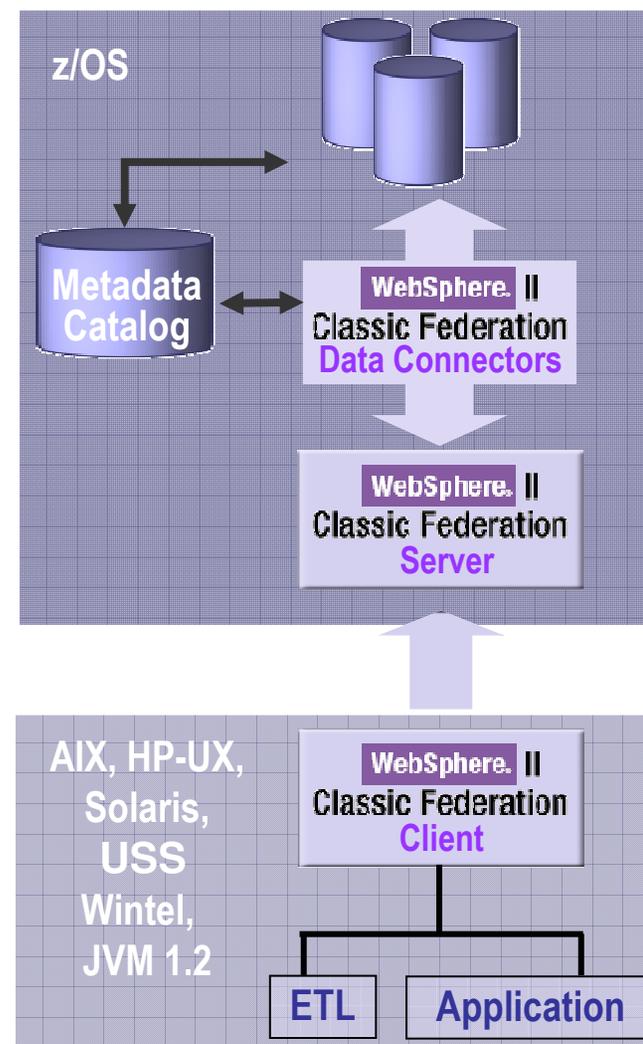
- **Inconvénients :**

- ▶ Les paramètres USS doivent être ajustés pour tenir compte d'une exécution parallèle

Fédération – Accès temps réel

Classic Federation

- Lecture / Ecriture des données mainframes via SQL à l'aide de pilotes JDBC ou ODBC
- Des connecteurs sur le mainframe transcrivent les ordres SQL en accès natifs
- Une vue relationnelle des données est créée par mapping des définitions physiques des données.
- **Bénéfices**
 - ▶ Accès instantané aux sources de données
 - ▶ Pas de programmation mainframe nécessaire
 - ▶ Accès consolidé à toutes les sources de données au travers d'une seule interface standard
 - ▶ Pas de modification de l'application si de nouvelles sources de données sont ajoutées



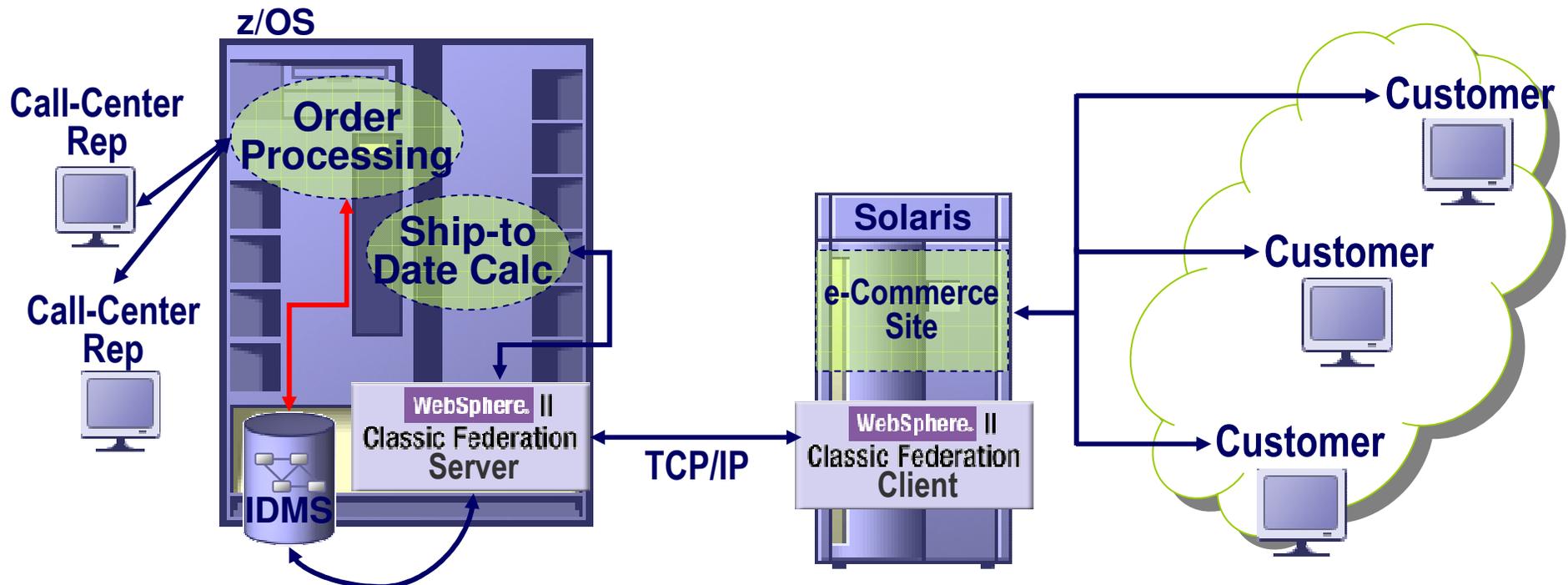
Cas Client : Site de commerce electronique

neckermann
SHOPPING



Partage transparent des données et processus critiques

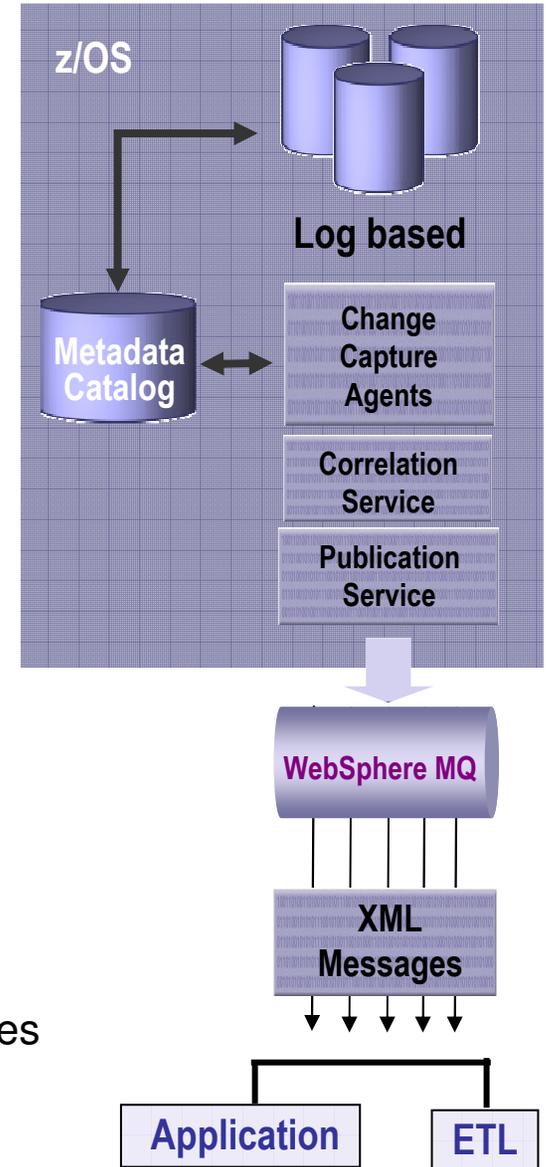
- ▶ Les applications e-commerce partagent des données critiques à jour à chaque instant
- ▶ Réutilisation transparente des procédures de calcul communes (ship-to-date, pricing)
- ▶ Les équipes de développement WebSphere n'ont pas besoin de connaissances mainframe
- ▶ L'application e-commerce est modifiée sans impact sur les applications mainframe de call-centers



Publication – Fil de l'eau

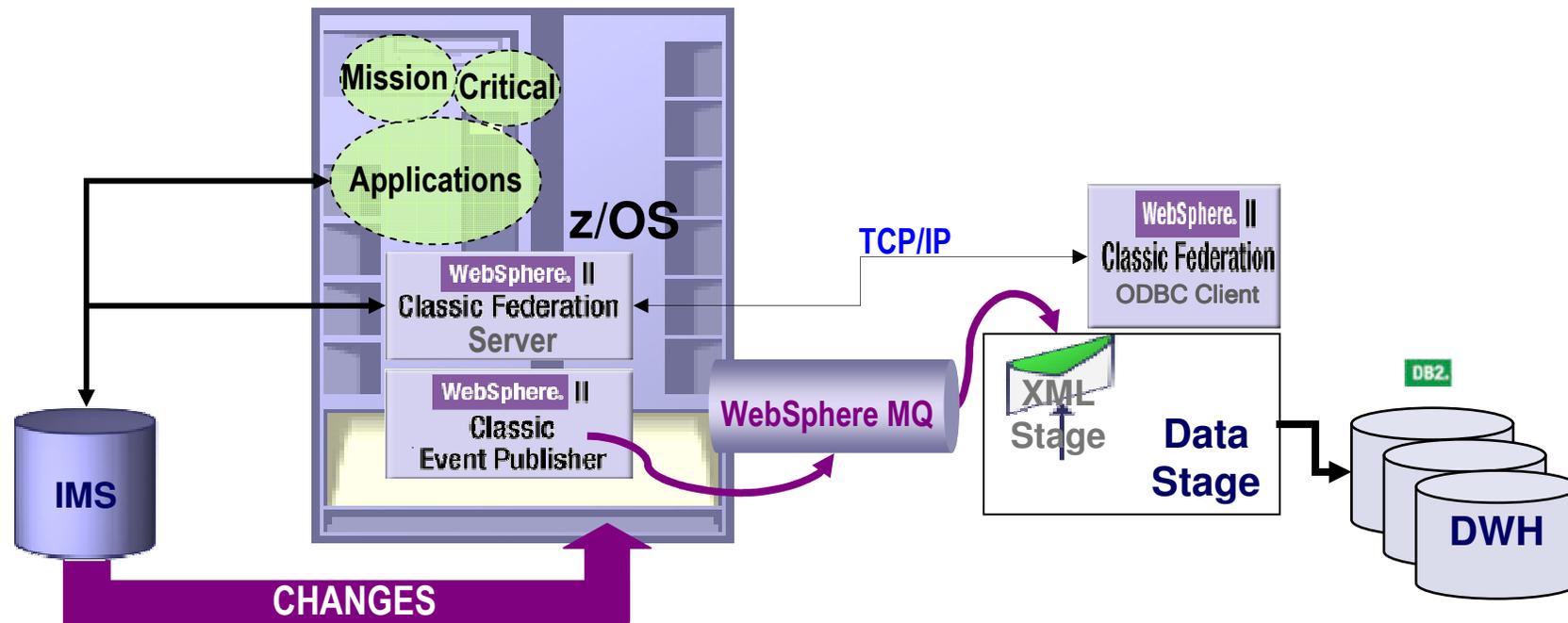
Classic Event Publishing

- Des agents capturent les modifications effectuées sur les sources de données et les envoient au service de Correlation.
- Le *service de corrélation* convertit les données capturées dans un format relationnel et conserve les mises à jour en attendant un ordre de COMMIT ou ROLLBACK.
- Le *service de publication* crée un message XML à partir des mises à jour validées, publie le message dans une queue WebSphere MQ et avertit le service de corrélation de la bonne livraison des données.
- **Bénéfices**
 - ▶ Alimentation au fil de l'eau des seules modifications effectuées
 - Réduit le délai de mise à disposition des informations
 - Minimise les ressources nécessaires à la capture des données
 - Optimise la bande passante Réseau
 - ▶ Assure une livraison fiable des données



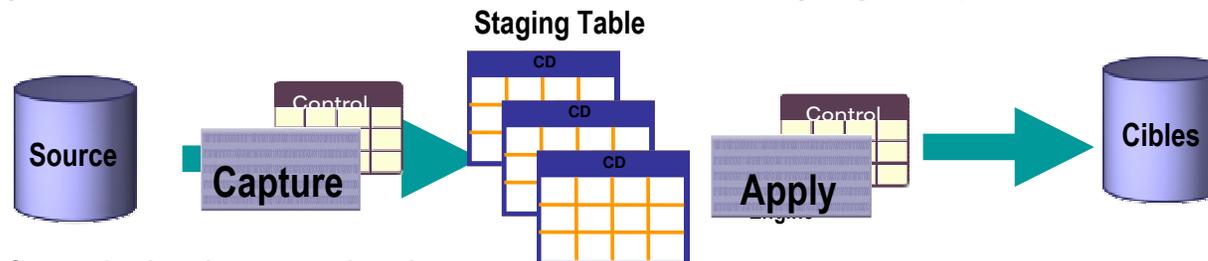
Cas Client : Alimentation d'un datawarehouse DB2 Banque Importante en Turquie

- Classic Federation et DataStage collaborent pour alimenter le datawarehouse à partir de données IMS (Initialisation et Mises à jour en masse).
- EventPublisher et DataStage sont utilisés pour une mise à jour au fil de l'eau.
- Aucun champ n'était disponible pour identifier les données modifiées.



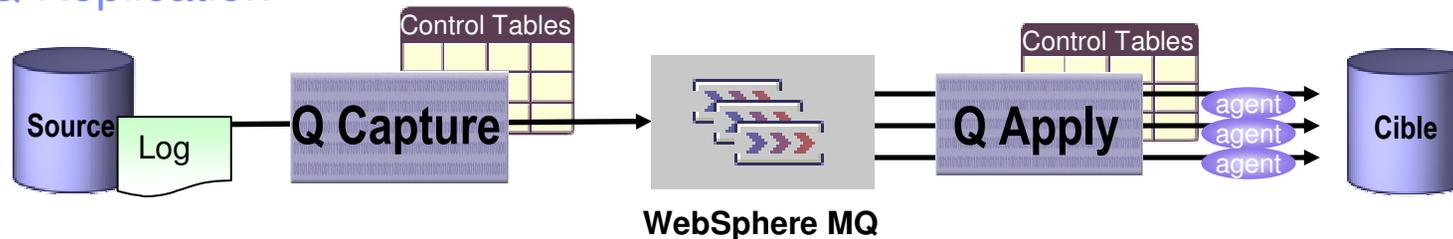
Réplication – Disponibilité continue

➤ SQL Replication (anciennement IBM DataPropagator)



- Capture à partir des logs ou de triggers
- Utilisation de tables intermédiaires
- Capacité de transformation 1 vers N
- Grande Flexibilité d'ordonnancement et de distribution des données

➤ Q Replication



- Capture à partir des logs et transport via WS MQ
- Application des mises à jour en parallèle
- Adapté au contexte de forte volumétrie avec temps de latence faible
- Réplication multi-directionnelle (exemple : serveur hot-standby ou répartition de charges)
- Détection et Résolution automatique des conflits
- Sources / Cibles de structures similaires

Cas Client : Disponibilité d'un service Client

Fournisseur international de services financiers

Challenge

- Fournir aux clients de meilleur temps de réponse lors de requêtes de interactives.
- Répartition des données critiques sur plusieurs sites.

Solution

- Q Replication de plus de 10 Million de transactions vers un site secondaire distants de plusieurs milliers de kilomètres. Le mécanisme de réplication est aujourd'hui mono-directionnel. Il est prévu de passer à un mode d'égal à égal.

Bénéfices

- Temps de latence inférieur à 2 s.
- Une utilisation plus efficace et moins coûteuse des ressources
- Le deuxième site assure des services de reporting et d'analyse tout en jouant le rôle de site de sauvegarde.
- Une réplication en temps réel du deuxième site apporterait comme autre bénéfice une capacité de traitement supplémentaire en cas de pics de charge.

Agenda

- Contexte et Enjeux
- Modèles d'intégration
- **Cas Client**
- Synthèse

Waitrose.com

*Filiale de John Lewis Partnership
Acteur majeur du commerce
de détail en Angleterre
174 supermarchés
CA 2006 : £ 3 Milliard
36 651 employés*



Environment technique

- Fort existant Mainframe
 - Data Warehouse sur DB2 z/OS
 - Serveur Web sous Linux z/VM
 - DataStage USS & RedHat Linux
- Configuration matérielle
 - *z990 2084-B16 avec 8 CPs ('model 308'), 1 ICF and 4 IFLs*
 - *DataStage LPAR : 4 logical CPs, 135 MSU, 10 GB storage*

Processus applicatifs

- *Construction d'états de synthèse quotidiens*
- *3 millions de lignes - 17 mins, 38 000 lignes - 10 mins, 197 lignes - 2 mins*
- 50 traitements, En production depuis 12 mois

Cas Client : Refonte du système d'alimentation BI sur system Z



■ Challenge

- ▶ Choix d'entreprise de placer le datawarehouse sous DB2 z/OS
- ▶ 80 % des sources de données sont de type mainframe (DB2 z/OS, IMS, VSAM)
- ▶ 20 % des sources de données sont sous SAP, SQL Server, Oracle ou Excel.
- ▶ Toutes les chaînes d'alimentation étaient des développements spécifiques
- ▶ Volonté de déployer une technologie ETL pour améliorer la productivité des équipes

■ Solution : DataStage USS et Redhat Linux

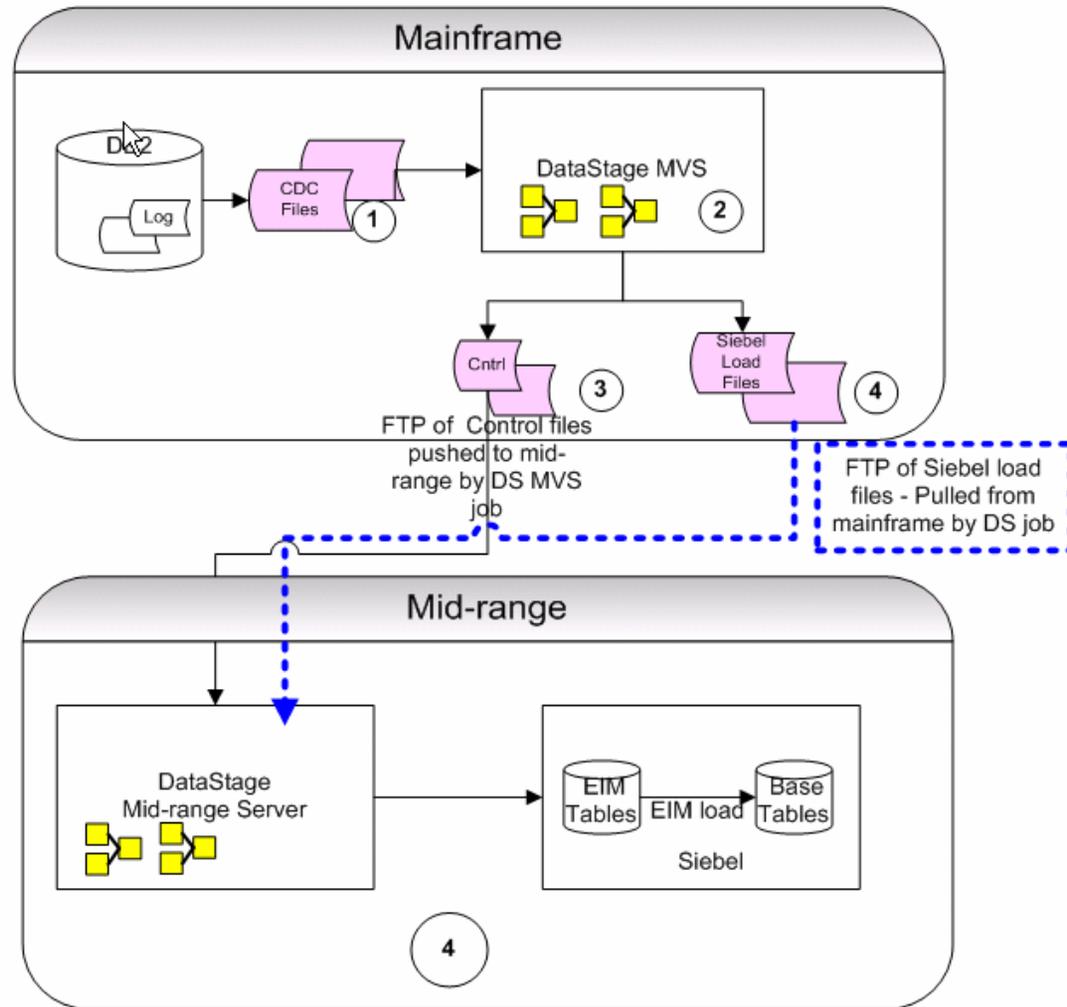
■ Bénéfices :

- ▶ **Réduction importante des coûts et des délais de développement** : une seule interface de développement, Composants génériques, Réutilisation accrue.
- ▶ **Reduction des coûts d'administration et de maintenance** : un seul référentiel.
- ▶ **Optimisation des compétences**
 - Un seul outil à maîtriser Unix – z/OS
 - Connaissance Unix valide aussi bien sur Z que sous Linux

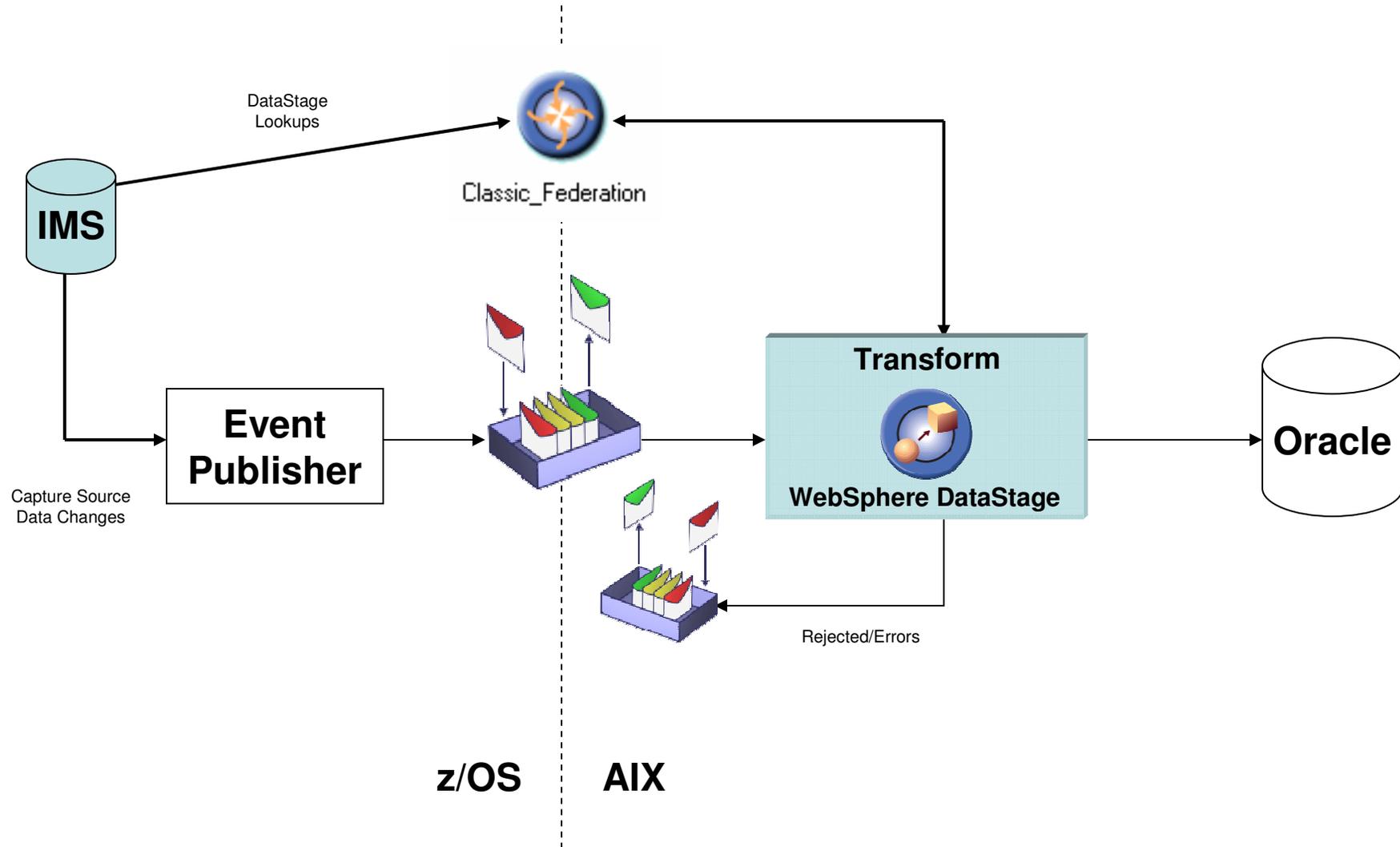
Agenda

- Contexte et Enjeux
- Modèles d'intégration
- Cas Client
- **Synthèse**

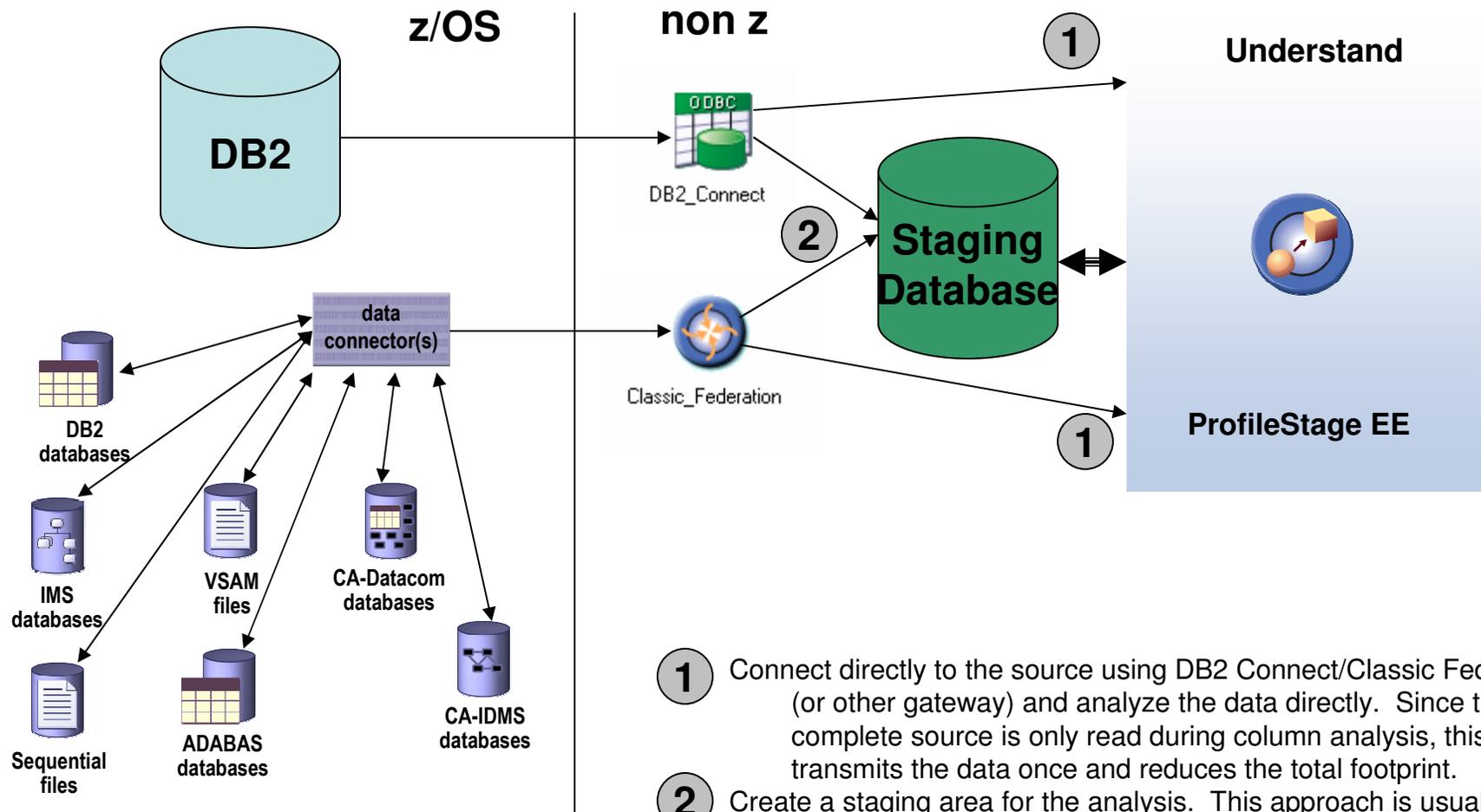
Architecture Transformation – Génération Code COBOL



Architecture Transformation – Moteur de transformation



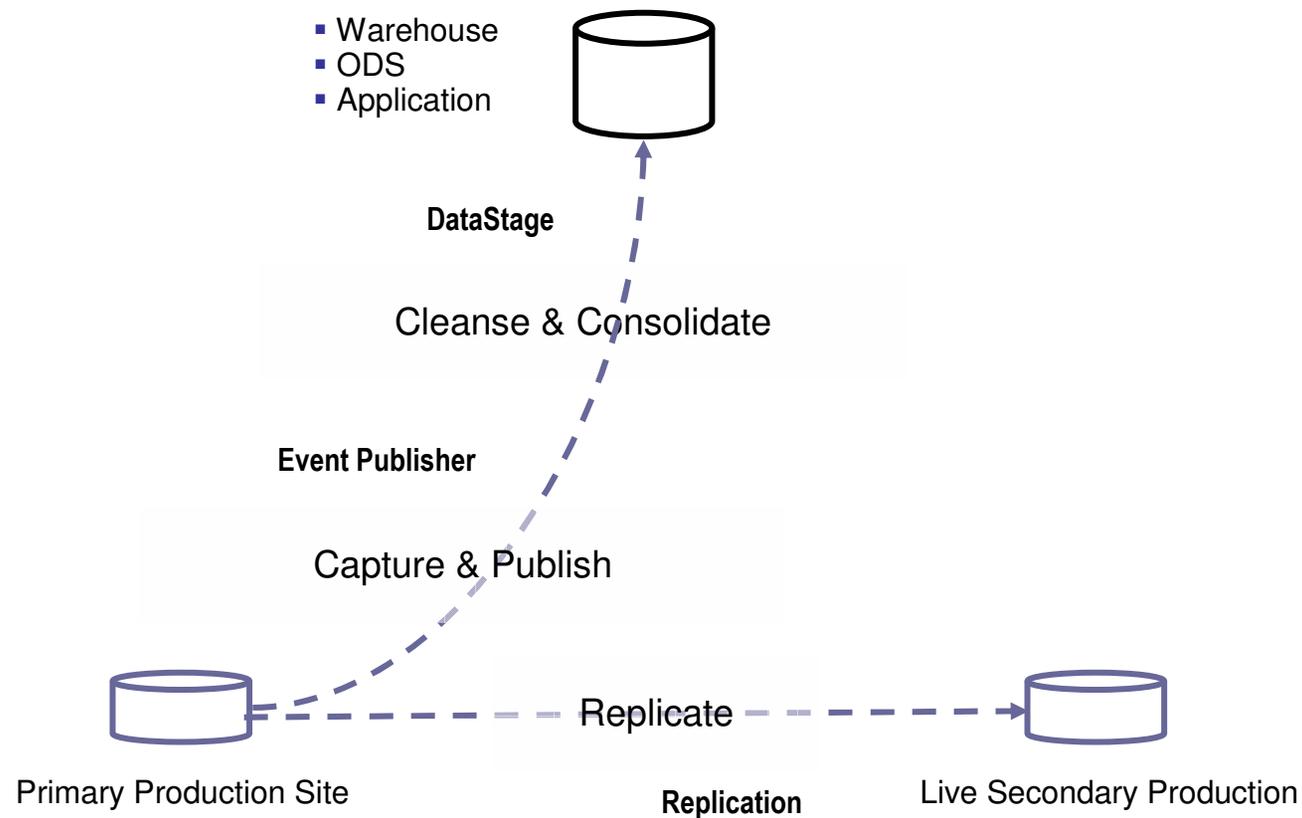
Architecture Découverte de données



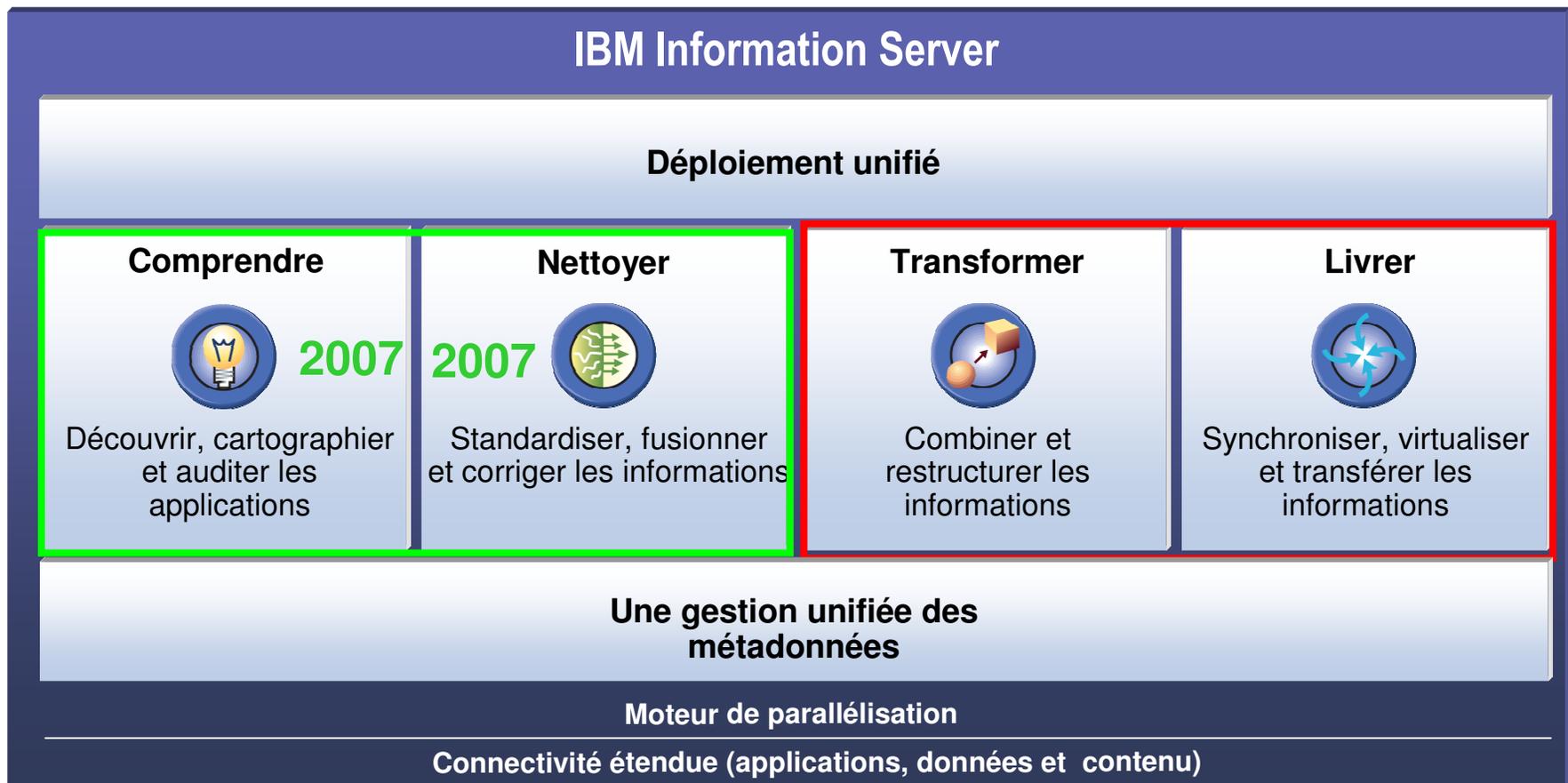
- 1 Connect directly to the source using DB2 Connect/Classic Federation (or other gateway) and analyze the data directly. Since the complete source is only read during column analysis, this only transmits the data once and reduces the total footprint.
- 2 Create a staging area for the analysis. This approach is usually applied to a statistically valid subset of the data. This avoids the “twinkling data” phenomena if analysis has to be redone and can provide efficiencies for iterative analysis. It also isolates the effect of users drilling into the data.

Supériorité d'une plate-forme complète d'intégration de données : Combiner les modèles.

Utiliser l'infrastructure Information Intergration pour publier vers le datawarehouse tout en maintenant une sauvegarde du système de production



Tout IBM Information Server sur system Z en 2007



Pour plus d'informations

- Gautier Rose
 - gautier.rose@fr.ibm.com