



# IBM InfoSphere : Solution de gestion de l'information

**Aomar BARIZ** [aomar.bariz@fr.ibm.com](mailto:aomar.bariz@fr.ibm.com)  
IBM Technical Sales Specialist Data Integration



**ON DEMAND BUSINESS™**

© IBM Corporation

# Agenda

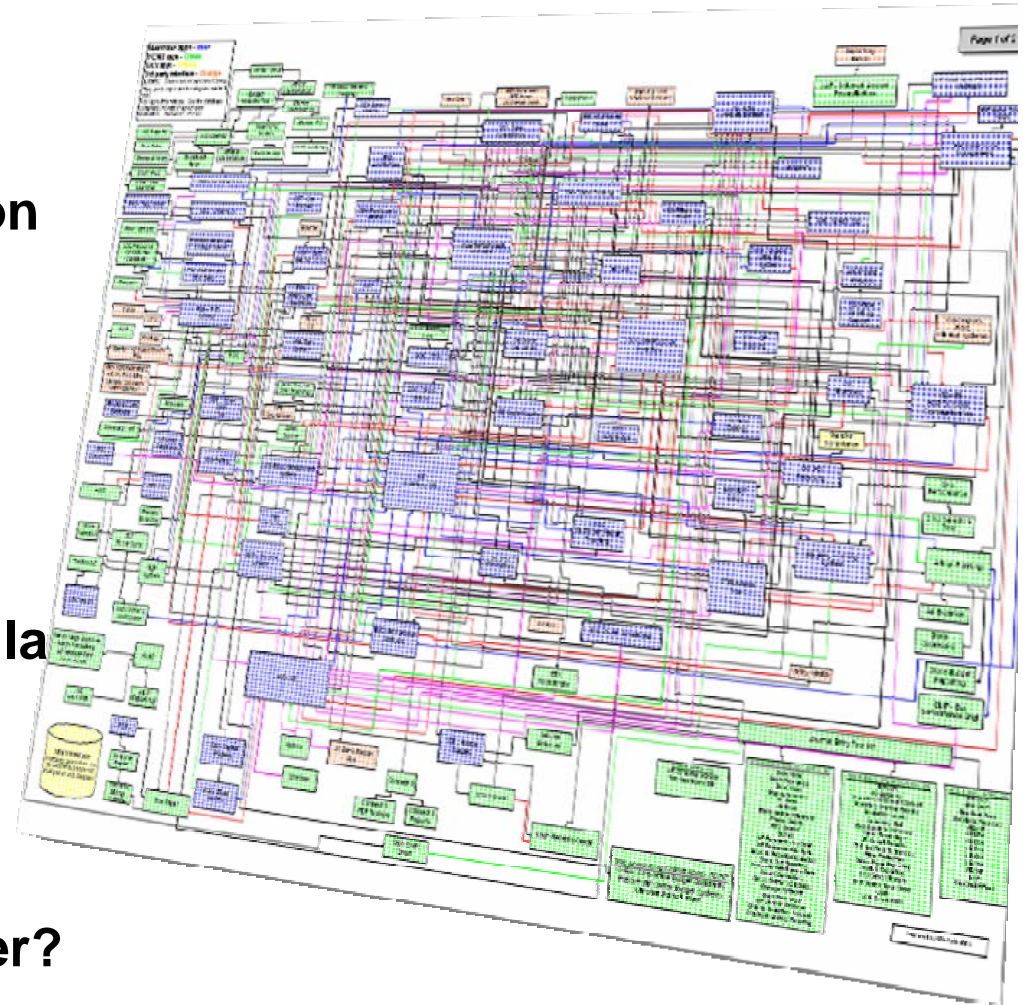
- Introduction : enjeux de l'intégration de données
  
- Présentation Information Server
  - Qualité des données : *Intégration rapide de nouvelles données, reprise de système, dédoublonnage de référentiels*
  - Transformer : *DataStage Enterprise Edition, Nouveautés de la version 8 (disponible depuis octobre 2006)*
  - Accélérer vos projets : *Génération de flux DataStage à partir de spécifications fonctionnelles*
  - Réduire les fenêtres batch et créer un entrepôt de données dynamique : *Capture des données modifiées en temps réel et diffusion vers un système cible*
  - Gestion des méta données : *Metadata Server/workbench: Maintenance des flux, Analyse d'impact, Suivi de règles métier, Modification de structure reporting Cognos*
  
  - Services d'information : *Publication de services de transformation, intégration dans une architecture orientée services"*
  
- IBM InfoSphere Roadmap
  
- Conclusion et apport Information Server

# Agenda

- Introduction : enjeux de l'intégration de données
  
- Présentation Information Server
  - Qualité des données : *Intégration rapide de nouvelles données, reprise de système, dédoublonnage de référentiels*
  - Transformer : *DataStage Enterprise Edition, Nouveautés de la version 8 (disponible depuis octobre 2006)*
  - Accélérer vos projets : *Génération de flux DataStage à partir de spécifications fonctionnelles*
  - Réduire les fenêtres batch et créer un entrepôt de données dynamique : *Capture des données modifiées en temps réel et diffusion vers un système cible*
  - Gestion des méta données : *Metadata Server/workbench: Maintenance des flux, Analyse d'impact, Suivi de règles métier, Modification de structure reporting Cognos*
  
  - Services d'information : *Publication de services de transformation, intégration dans une architecture orientée services"*
  
- IBM InfoSphere Roadmap
  
- Conclusion et apport d' Information Server

## Les challenges ...

- Où est l'information ?
- Comment la récupérer au bon moment ?
- Quelle est sa signification ?
- Est-elle fiable ?
- Comment la récupérer dans la forme requise ?
- Comment la contrôler ?
- Comment la corriger/nettoyer ?
- ...

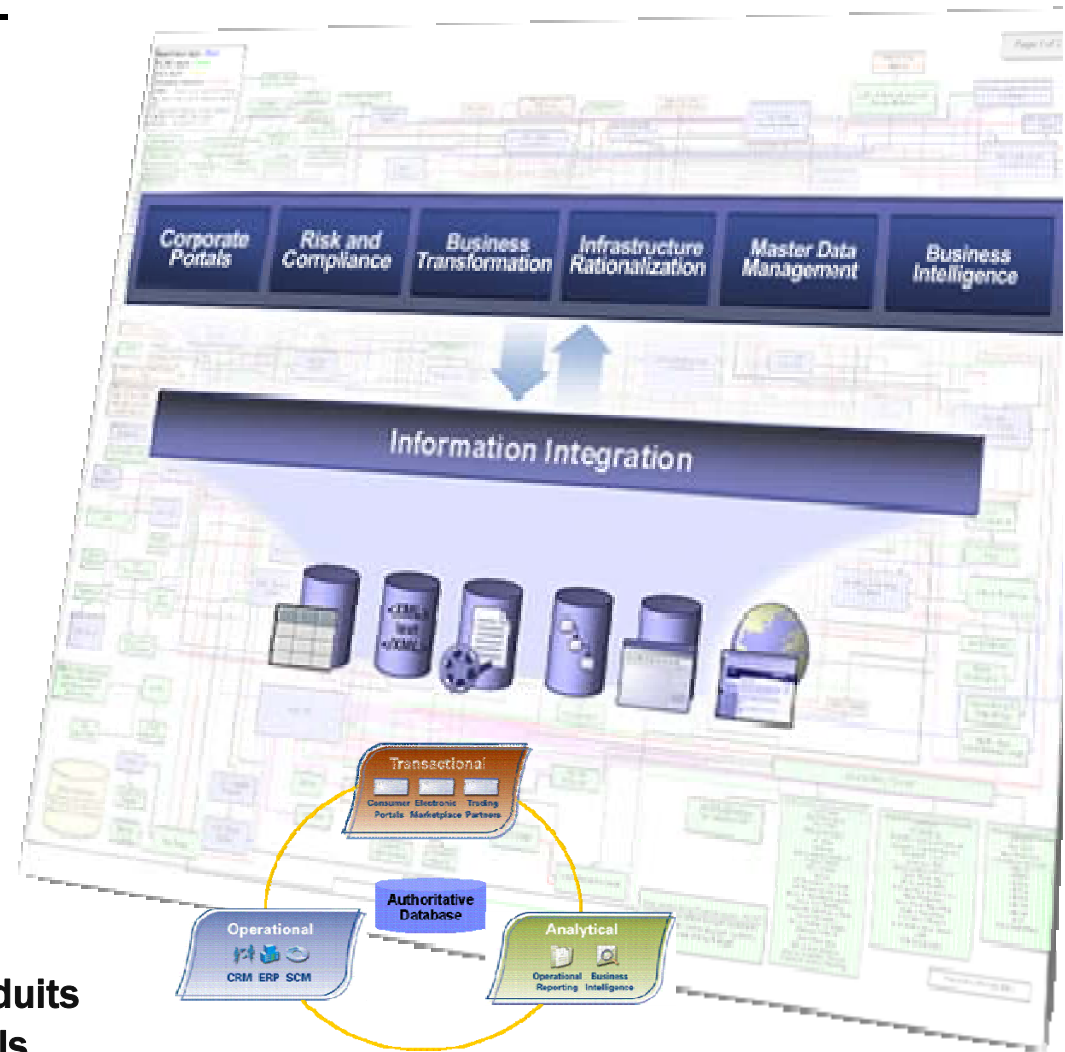




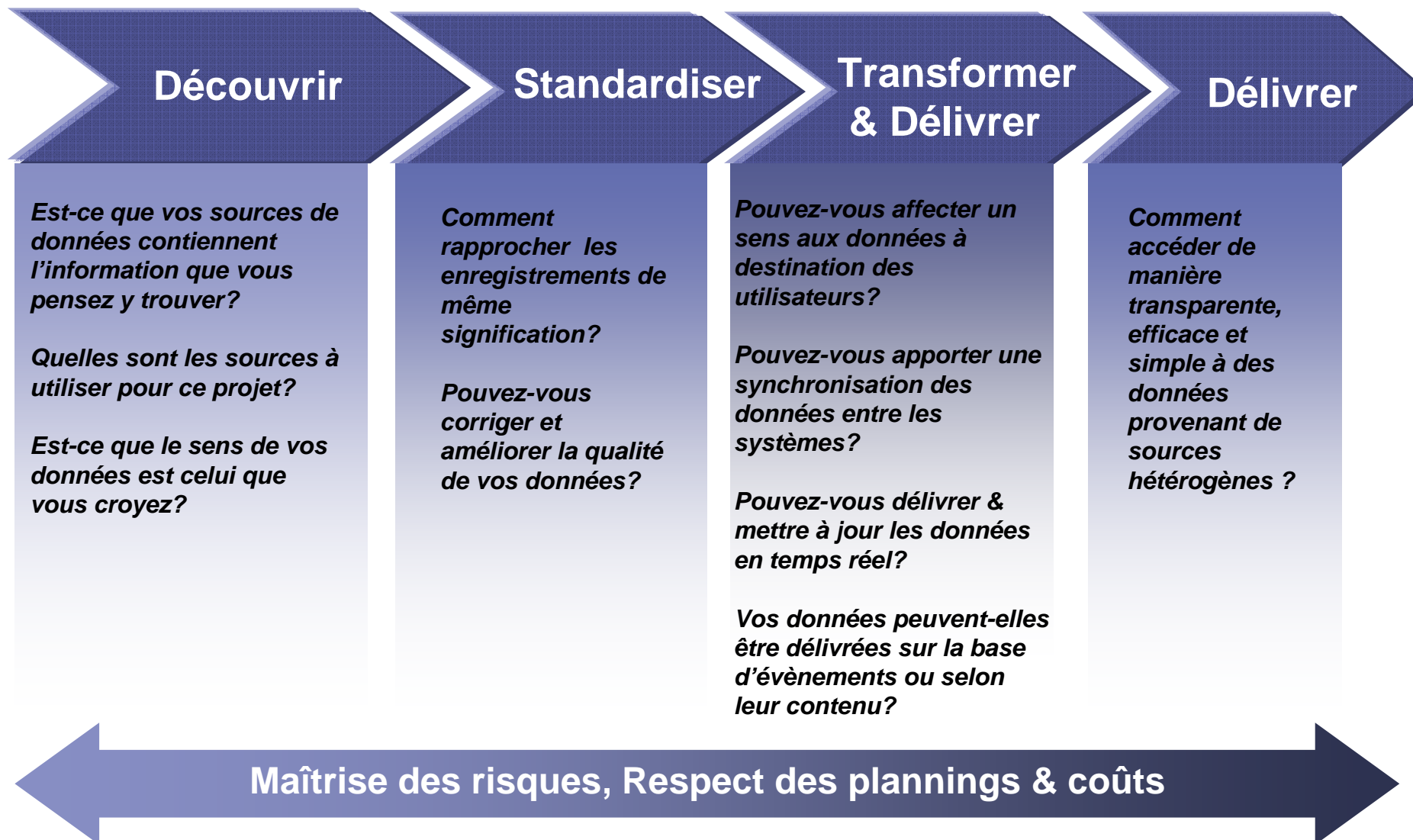
## L'économie de l'information

- Les systèmes dits 'mission-critical' sont maintenant centrés sur l'information
- Gérer l'intégration de l'information par projet est coûteux
- Seule une infrastructure intégrée de gestion de l'information peut répondre à ces défis

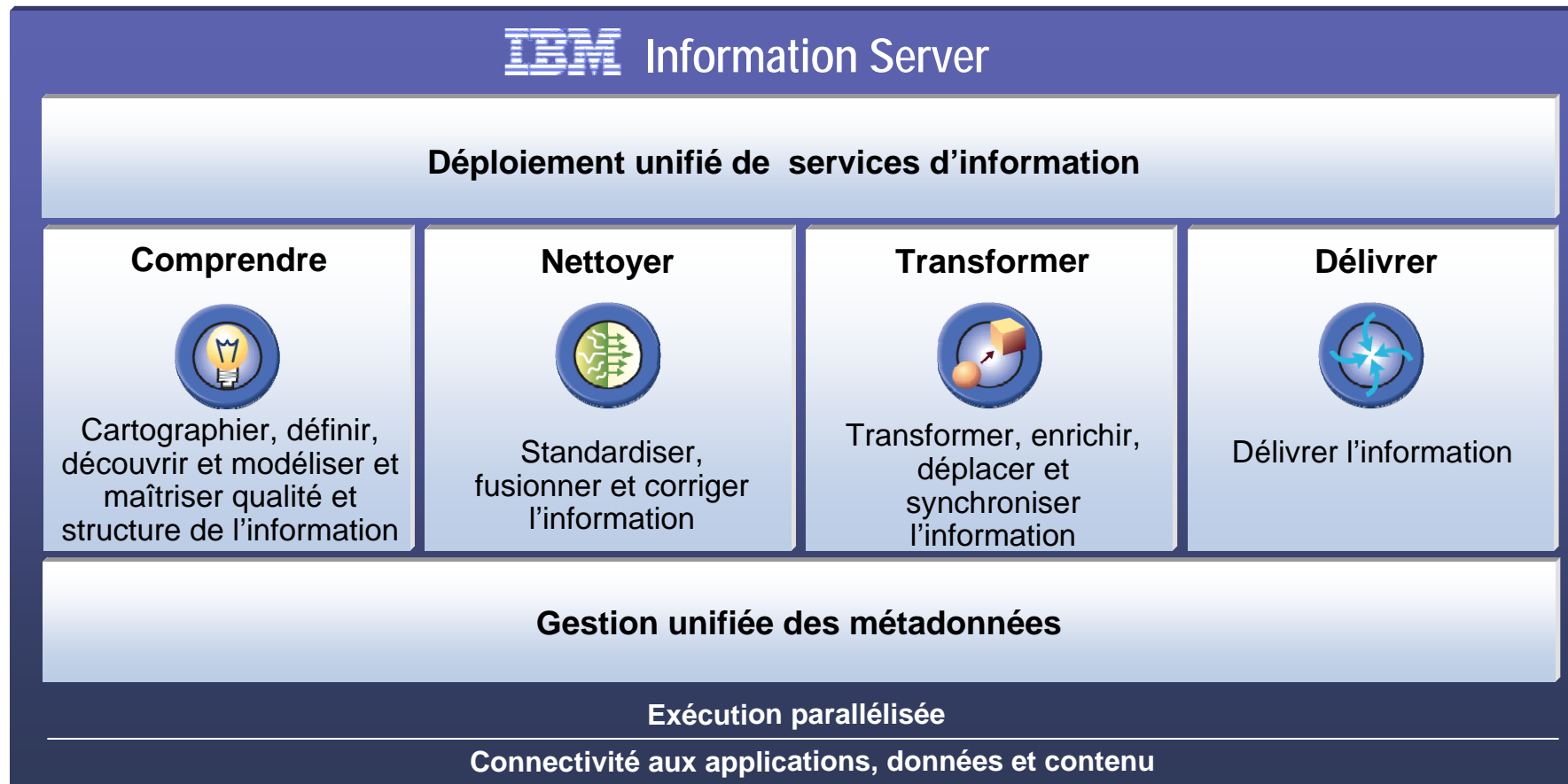
**Productivité**  
**Réutilisation et uniformité**  
**Coûts de développement inférieurs**  
**Coûts d'administration et de support réduits**  
**Alignement des rôles plurifonctionnels**



## Une méthodologie adaptée à ces enjeux métiers



# La plateforme IBM d'Intégration de données : IBM Information Server



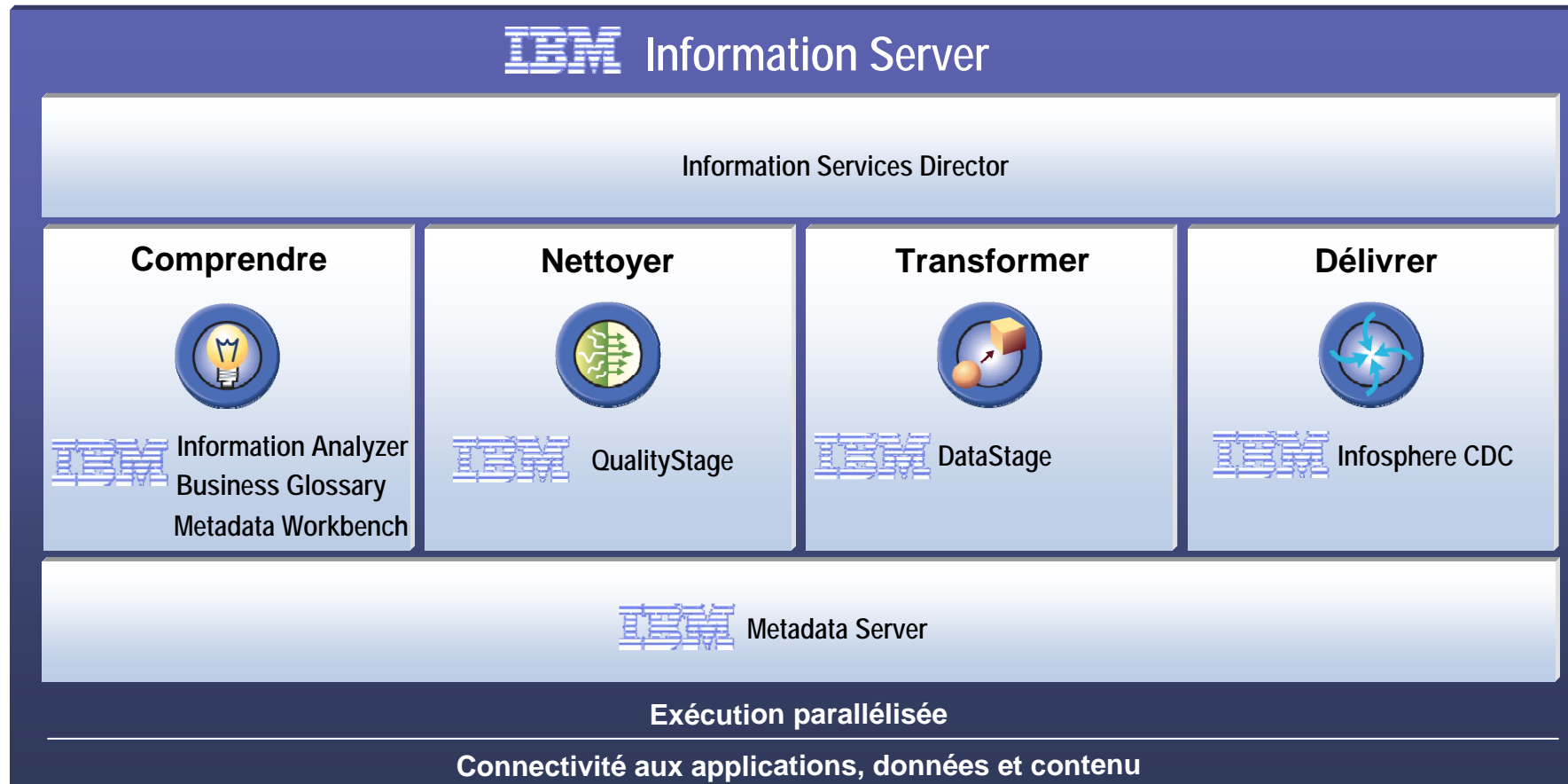
# Agenda

- Introduction : enjeux de l'intégration de données
- Présentation Information Server
  - Qualité des données : *Intégration rapide de nouvelles données, reprise de système, dédoublonnage de référentiels*
  - Transformer : *DataStage Enterprise Edition, Nouveautés de la version 8 (disponible depuis octobre 2006)*
  - Accélérer vos projets : *Génération de flux DataStage à partir de spécifications fonctionnelles*
  - Réduire les fenêtres batch et créer un entrepôt de données dynamique : *Capture des données modifiées en temps réel et diffusion vers un système cible*
  - Gestion des méta données : *Metadata Server/workbench: Maintenance des flux, Analyse d'impact, Suivi de règles métier, Modification de structure reporting Cognos*
  - Services d'information : *Publication de services de transformation, intégration dans une architecture orientée services"*
- IBM Infosphere Roadmap

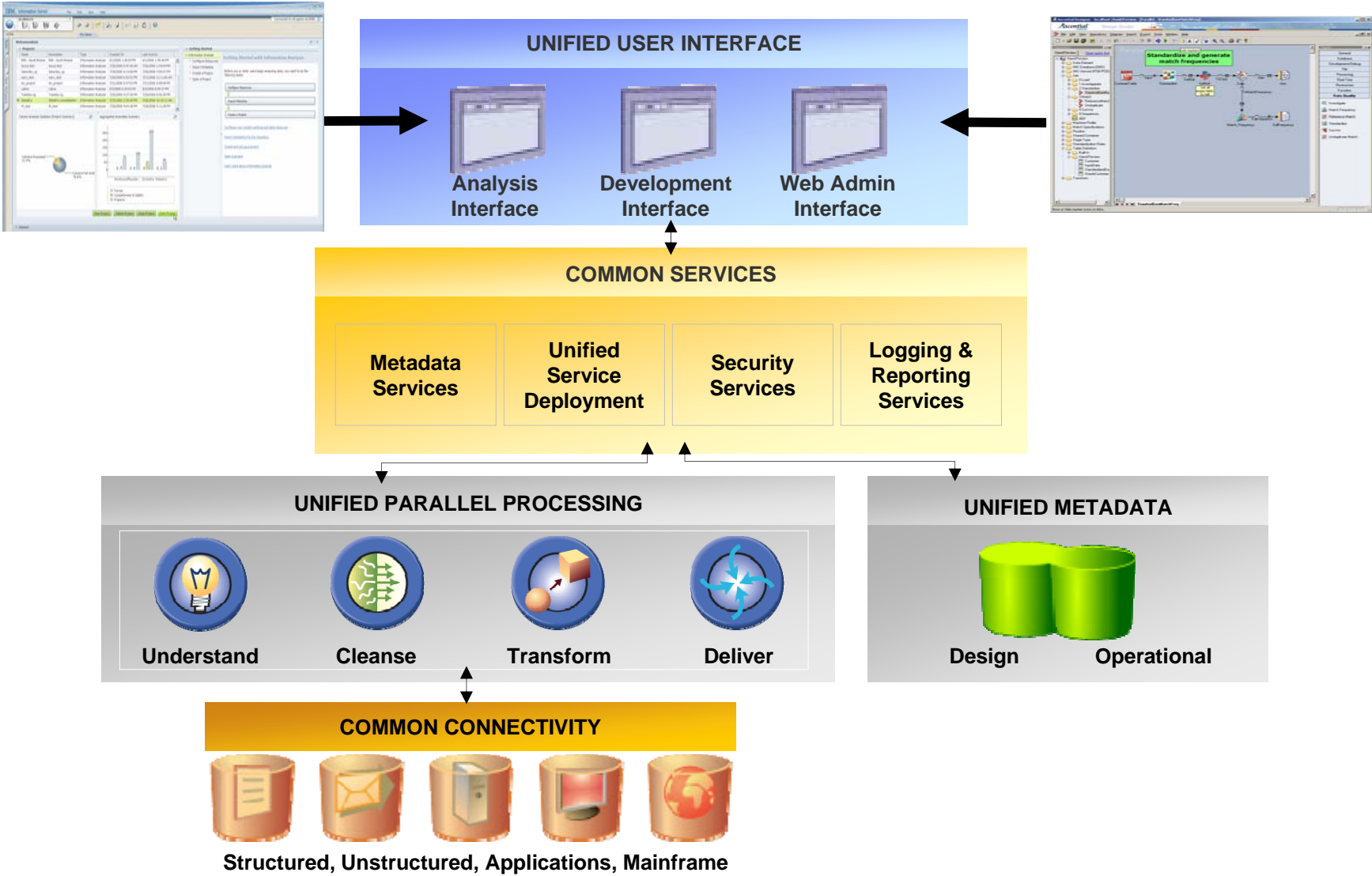


# IBM Information Server

*“Delivering information you can trust”*

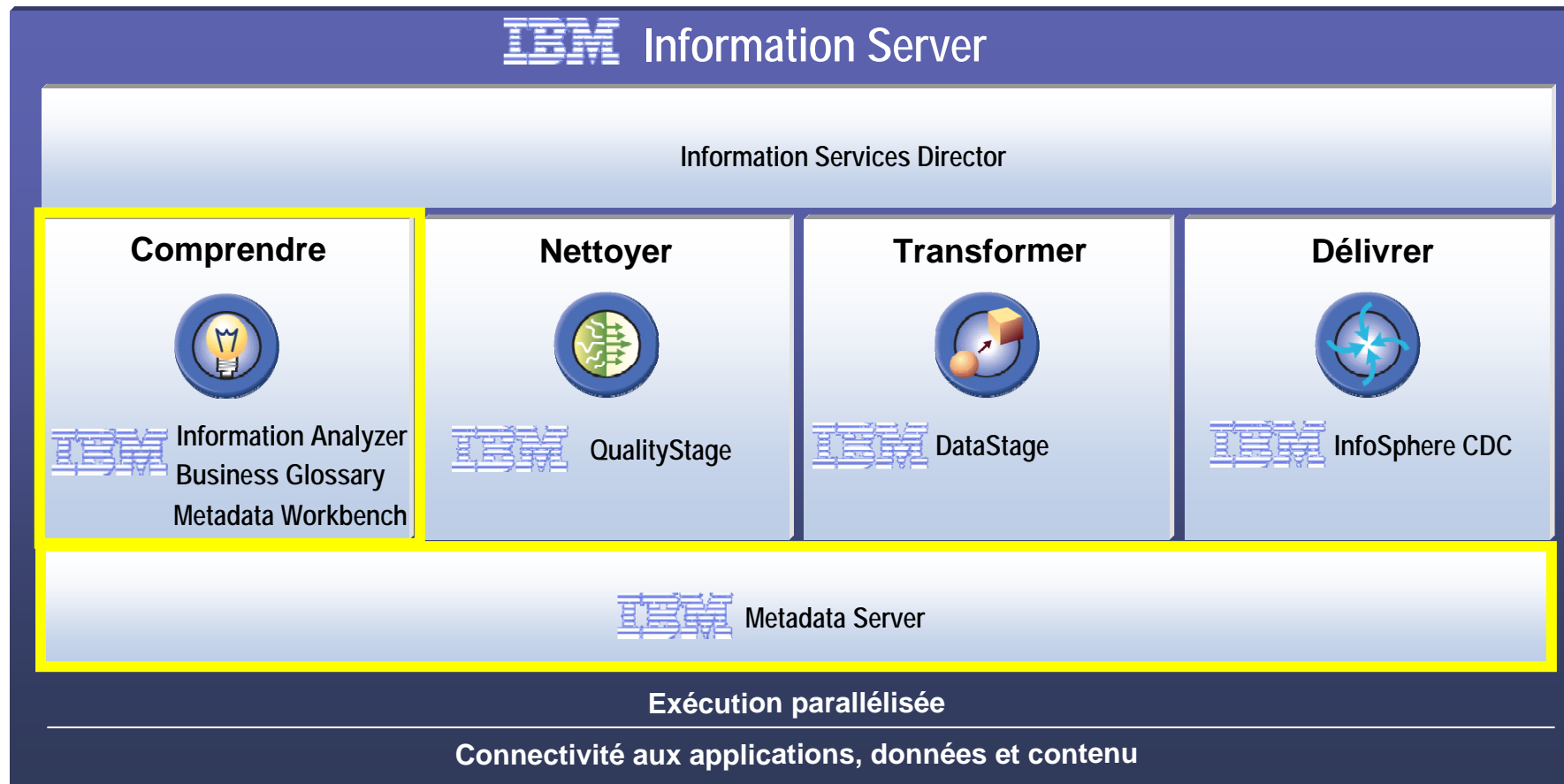


# Une Architecture unifiée et partagée

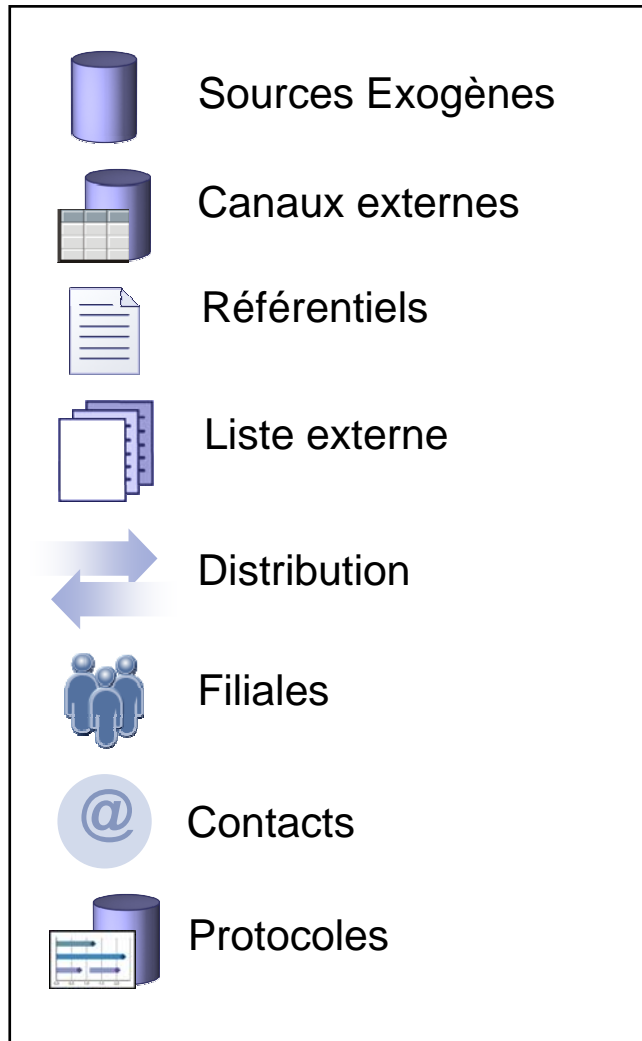


# IBM Information Server

*Delivering information you can trust*



# Comprendre



## Problèmes critiques :

- **Connaissez-vous réellement les données de votre SI ?**
- **Vous avez certainement de nouvelles sources de données ou d'autres encore inconnues ...**

## Pourquoi ?

- **Vos données et leurs relations sont incohérentes et divergent par rapport aux règles métier**
- **La documentation, si elle existe est incomplète, obsolète ou pire ... fausse !**
- **Les sources de données ne sont jamais statiques ... elles évoluent sans prévenir.**

## Stratégies adoptées

- **Travail Manuel intensif**
- **Impossible de passer en revue tous les éléments**
  - travail souvent approximatif ...
- **Sans infrastructure : comment rejouer une analyse ?**
- **Manque d'approche standardisée**

# Service Découverte (Information Analyzer)

- Cartographie détaillée et exhaustive de contenu, structure et qualité des données existantes
- Contrôle conformité par rapport à un modèle, règle
- Permet une mesure continue et la production de documentation sur les données

## Analyse de colonnes

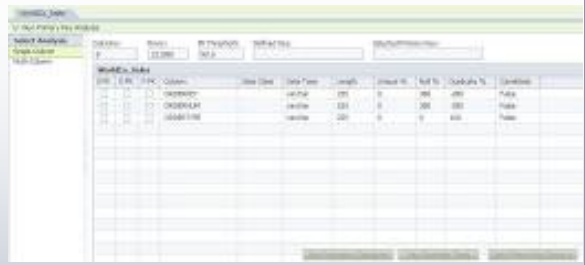


Fréquences de distribution



Analyse des Classe, propriétés, format, domaine/complétude

## Analyse de Table

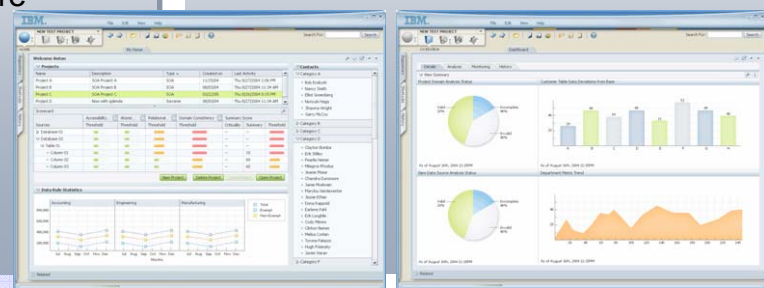


Analyse de clé primaire

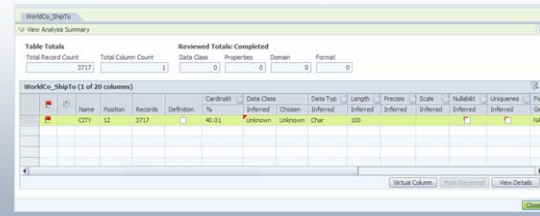
## Analyse Inter-Table



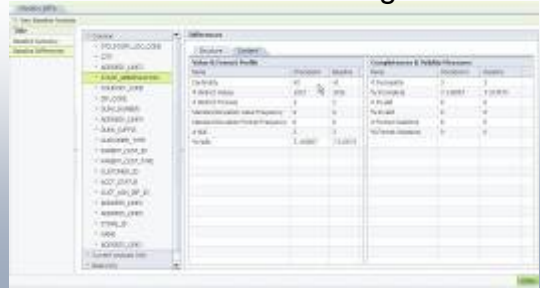
Analyse des dépendances



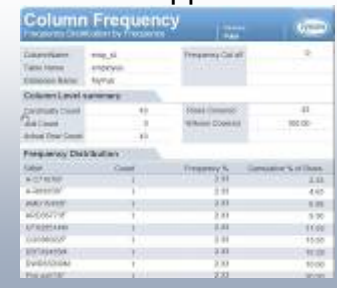
## Annotations & Marquage pour revue



## Validation de règles



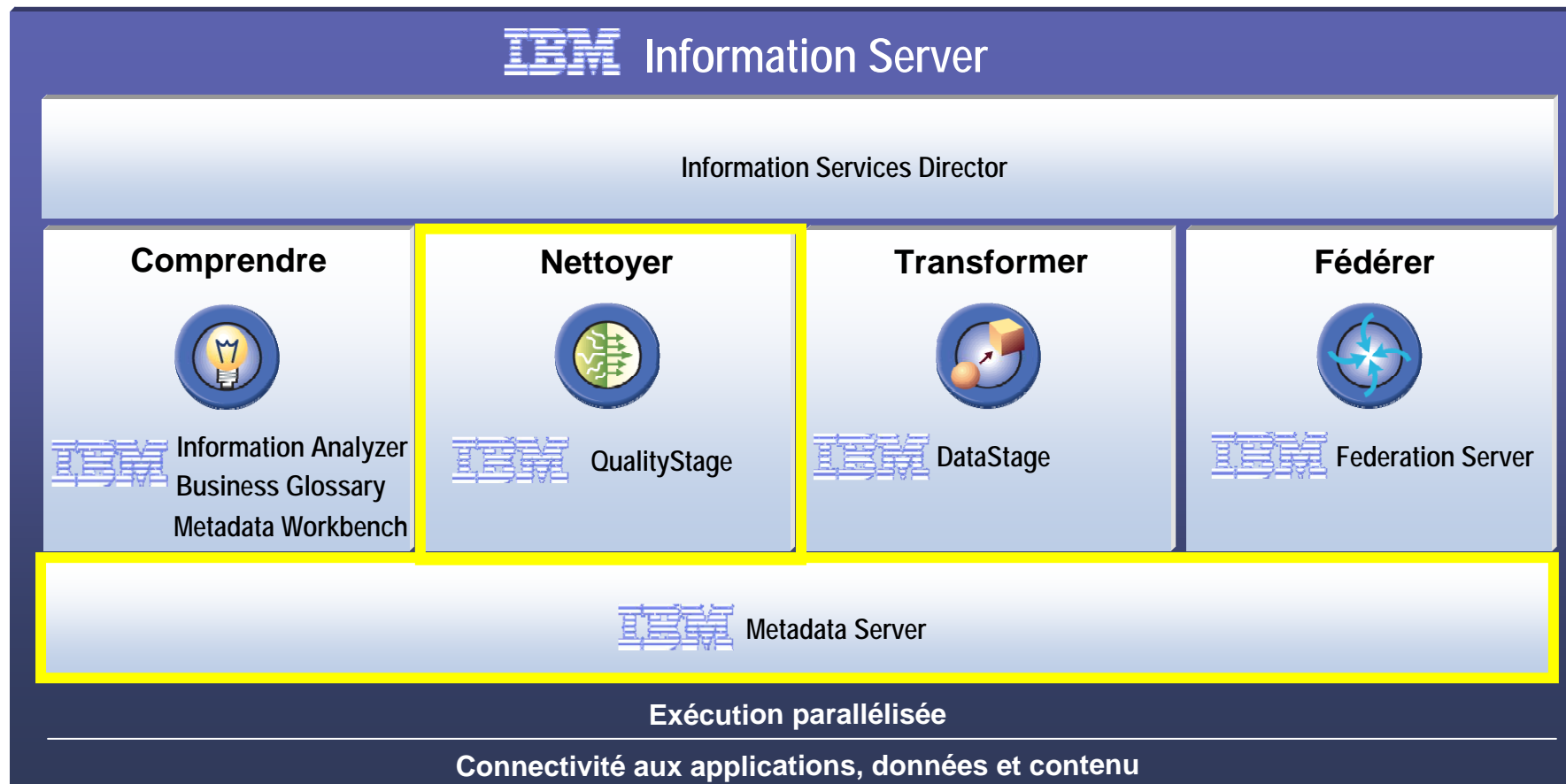
## Rapports






# IBM Information Server

*Delivering information you can trust*




# IBM QualityStage

- **Standardisation de champs en saisie libre**
  - Tout type de données : Noms, Titre, Adresses, Description Produits, ...
  - Algorithmes de Phonétisation
  - Standardisation et Validation postale prédéfinies pour plus de 200 pays
- **Rapprochement probabiliste**
  - Plus de 24 algorithmes prédéfinis
    - Comparaison exacte ou floue de chaînes de caractères
    - Distance, Pourcentage, Dates
  - Aucune limite dans le nombre de critères utilisés pour effectuer le rapprochement
  - Entièrement personnalisable
- **Consolidation des enregistrements**
  - Algorithmes prédéfinis et personnalisables
  - Gestion des références croisées
- **Déploiement comme service au sein de l'entreprise**




**Experts Sujets**



**Analystes**

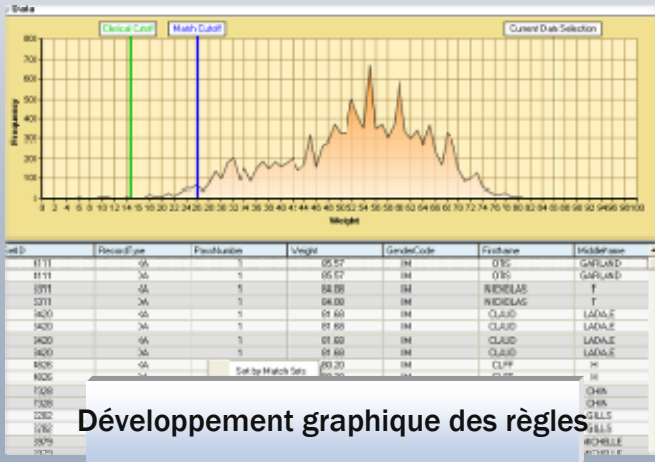
  



**Nettoyer**

**WebSphere QualityStage™**  
 Standardisation des données et mise en correspondance avec toutes les autres sources afin de définir une information unique

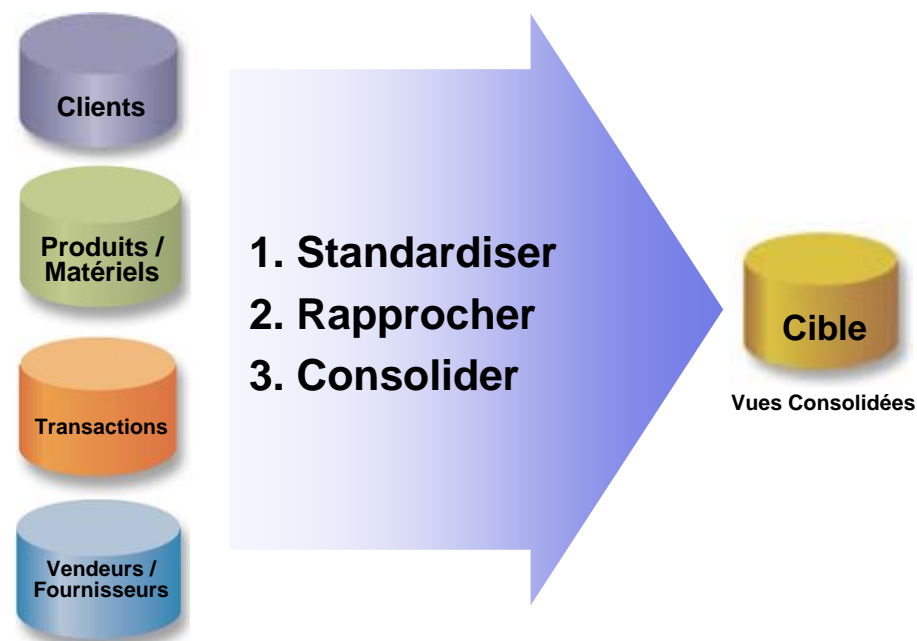
  



**Développement graphique des règles**

# Nettoyage des données avec IBM QualityStage *Processus*

- **Standardiser**
  - Décomposition basée sur les règles de standardisations
  - Analyse lexicale
  - Mise en contexte
  
- **Rapprocher**
  - Analyse basée sur des algorithmes probabilistiques
  - Identification des doublons
  - Identification des enregistrements de la “clerical zone”
  
- **Consolidation**
  - Constitution du “Best of Breed”
    - Information la plus fraîche
    - Information la plus complète



# IBM QualityStage : Un exemple d'harmonisation

## DONNEES EN ENTREE

Operation Work Instructions in a free text field

WNG ASSY DRL 3 HOLE USE HEXBOLT ¼ INCH  
 WING ASSEMBY, HEX BOLT .25" - DRILL FOUR, USE 5J868-A  
 USE 4 5J868A BOLTS (HEX .25) - DRILL HOLES FOR EACH ON WING ASSEM  
 RUDDER, TAP 6 WHOLES, SECURE W/KL2301 RIVETS (10 CM)

## STANDARDISATION

Assembly	Instruction	QTY	Type	Part	Size	Unit Measure	SKU
WING	DRILL	3	HOLES	HEXBOLT	.25	INCH	
WING	DRILL	4		HEXBOLT	.25	INCH	5J868A
WING	DRILL	4	HOLES	HEXBOLT	.25		5J868A
RUDDER	TAP	6	HOLES	RIVET	10	CM	KL2301

## RAPPROCHEMENT

Assembly	Instruction	QTY	Type	Part	Size	Unit Measure	SKU
WING	DRILL	3	HOLES	HEXBOLT	.25	INCH	
WING	DRILL	4		HEXBOLT	.25	INCH	5J868A
WING	DRILL	4	HOLES	HEXBOLT	.25		5J868A
RUDDER	TAP	6	HOLES	RIVET	10	CM	KL2301

MATCH

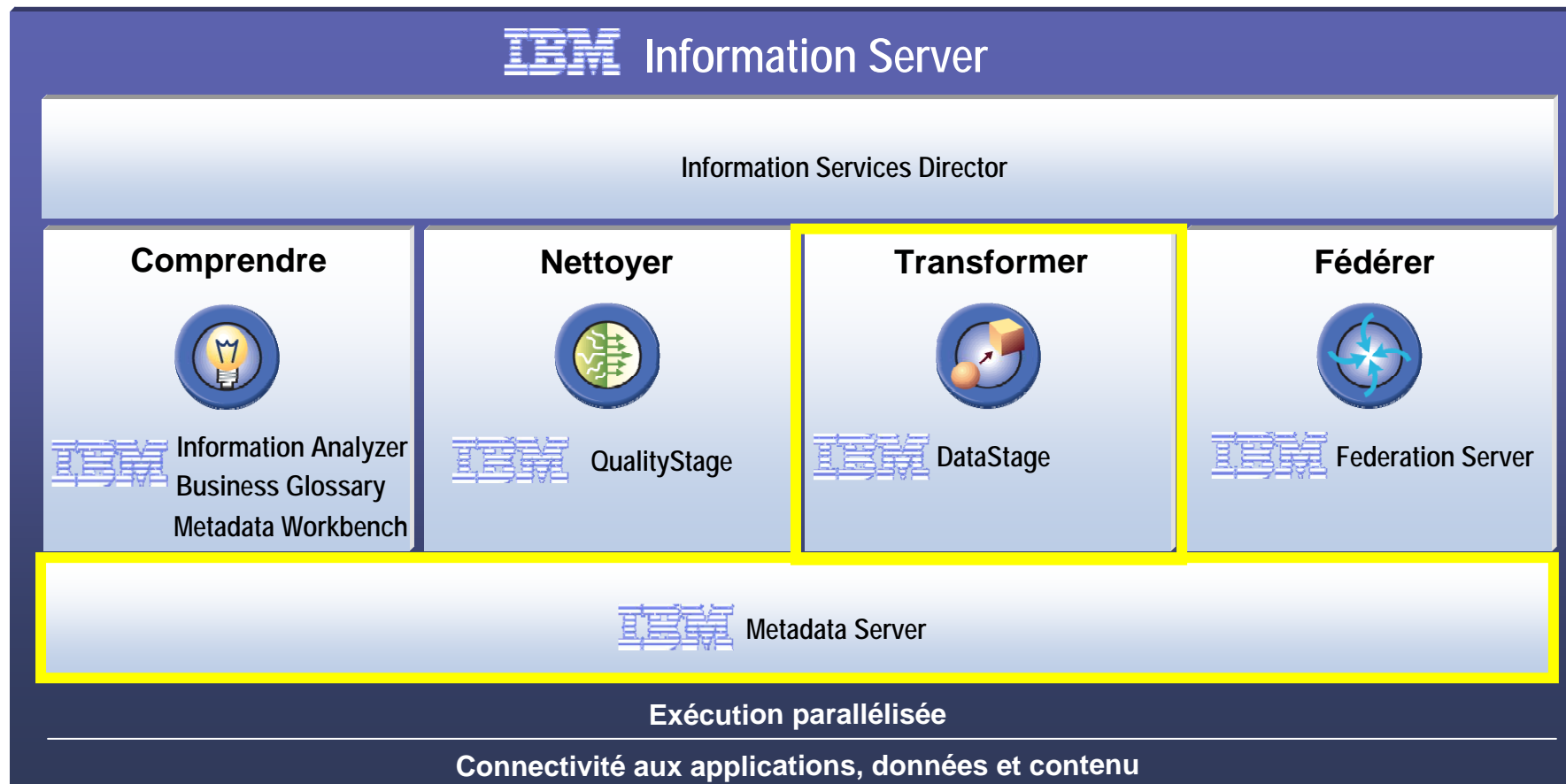
## CONSOLIDATION

Assembly	Instruction	QTY	Type	Part	Size	Unit Measure	SKU
WING	DRILL	4	HOLES	HEXBOLT	.25	INCH	5J868A
RUDDER	TAP	6	HOLES	RIVET	10	CM	KL2301



# IBM Information Server

*Delivering information you can trust*





# IBM DataStage Enterprise Edition (ETL)

Transformer



Développeurs



Architectes

- **Modélisation graphique des flux de données sans programmation**

- Plus de 50 composants génériques

- Connectivité : Fichier, XML, SGBDs, ERPs, Web Services, EJB, WS MQ, JMS, ...
- Transformation : Jointure, Tri, Agrégation, Lookup, Gestion des Dimensions changeantes, ...

- Plus de 300 fonctions de transformations

- Réutilisation de l'existant : librairies externes, procédures stockées, COBOL, SAS, ...

- **Support des architectures parallèles sans modification des développements**

- **Support du batch et du temps réel**

Transformer

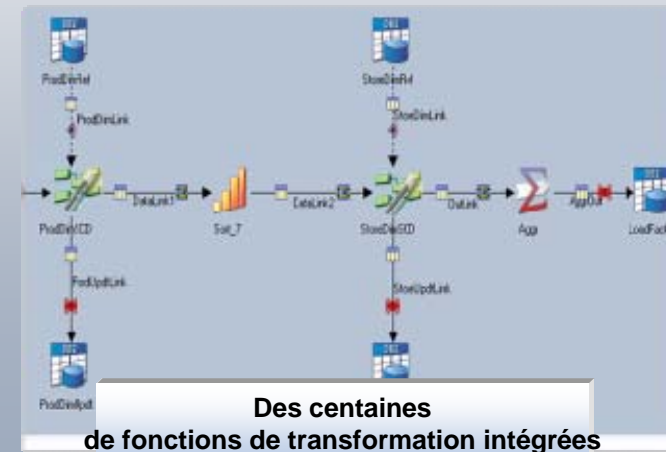


Livrer



IBM DataStage®

Transformer et agréger n'importe quel volume d'informations par lots ou en temps réel via une logique conçue de manière visuelle



## Développer - Modélisation du flux

- Modélisation graphique des flux de données et des processus

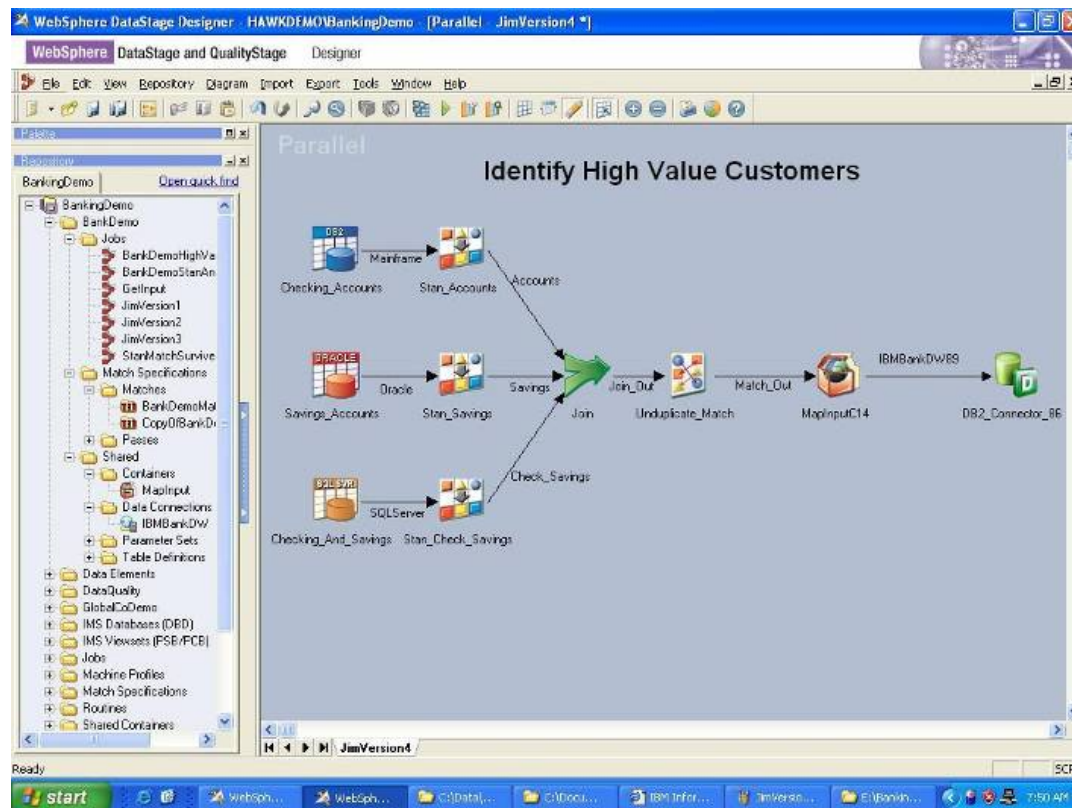
- Pas de programmation
- Compréhension facilitée
- Flexibilité totale
- Approche “composants”

- Méthodes lecture/écriture optimisées (native)

- Native, ODBC, Bulk Loaders
- XML, Mainframe, CDC, MQ,
- Applications , Web Services

- Multi-sources et multi-cibles en une seule passe

- Optimisation des ressources disques et CPU en adoptant une technique de partitionnement et parallélisme
- Gestion des *lookup* en mémoire



## Réutilisabilité / Ouverture

- **Routines** : création et référencement de bibliothèques de fonctions externes
- **Wrapper** : encapsulation et exécution parallèle de tout programme externe pour lire / écrire / transformer une donnée (ex. commande système grep, Programme SAS, outil de transfert CFT ou ConnectExpress, Calcul statistique, ...).
- **Buildop** : création de nouveaux composants graphiques prêts à l'emploi et répondant à une logique fonctionnelle complexe (exemple : agrégation conditionnelle) nécessitant une approche procédurale.
- **Connecteur Java** : interfaces avec toutes ressources Java (ex. JDBC, EJB, JMS, CA DATACOM).
- **Connecteur Web Services** : interfaces avec toutes ressources Web Services.

# Nouvelles fonctionnalités Designer

- Analyse d'impact graphique

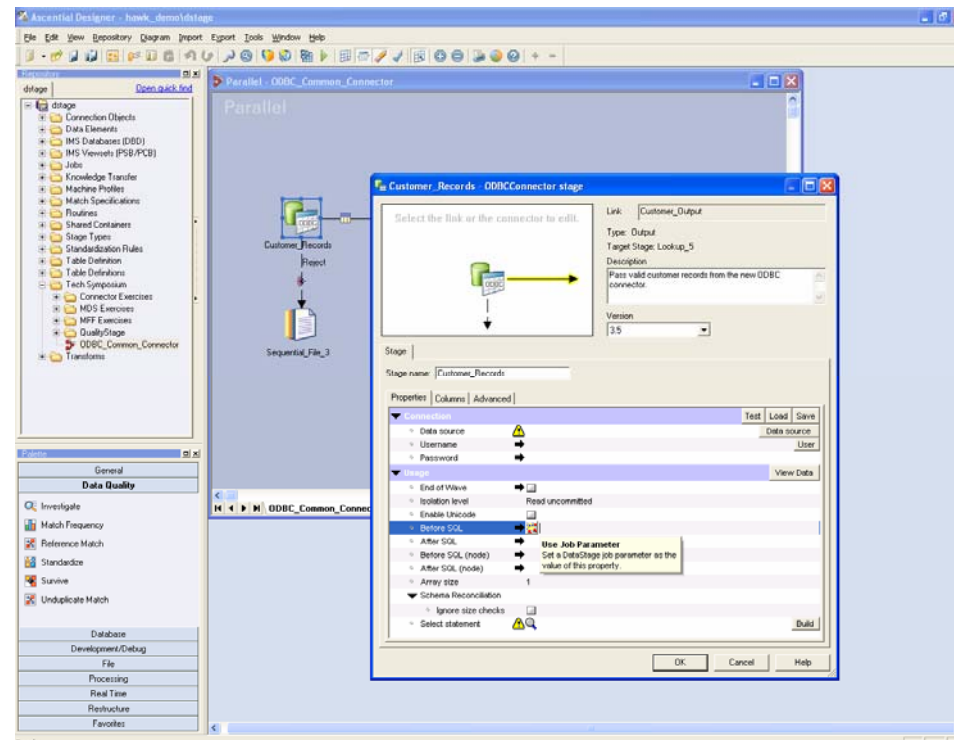
- Job, Table ou Routine Difference

```

Comparing Before_Table Against After_Table
- Properties (1 Change)
  Property Days Since Was Added As '2002-12-01!'
- Columns (6 Changes)
  Property Description Was Changed From Less Important Now To Important Field
  Property Nullable Was Changed From No To Yes
- Col4 (2 Changes)
  Property SQL Type Was Changed From LongVarChar To Date
  Property Length Was Removed
- Col9 (1 Change)
  Property Length Was Changed From 64 To 32
- Col10 (1 Change)
  Col10 Was Removed
    
```

## Les nouveautés en termes de connectivité

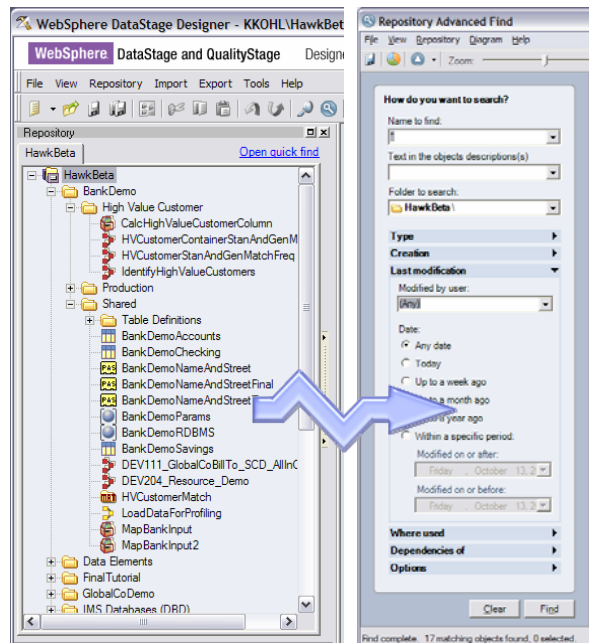
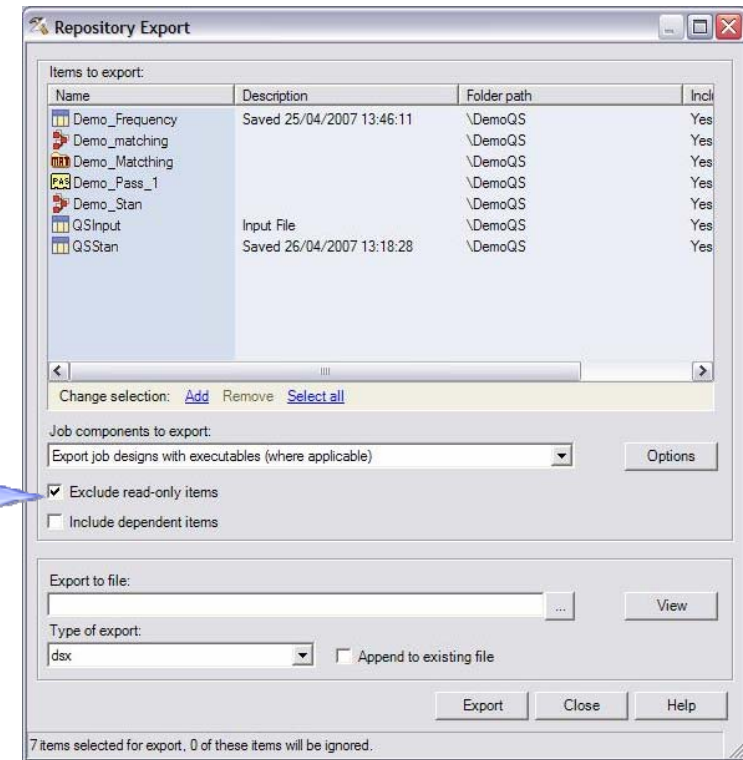
- Nouvelles fonctionnalités et SQLBuilder disponible pour :
  - DB2, Oracle, SQL Server, Teradata, ODBC, Sybase
- Procédures stockées support étendu pour:
  - DB2, Oracle, SQL Server, Teradata, Sybase
- Support versions SGBD (non exhaustif):
  - DB2 9.1, Oracle 10gR2, SQL Server 2005, Teradata v2r6.1 (DB server) / 8.1 (TTU), Sybase ASE 15, Sybase IQ 12.7
  - Informix 10 (IDS), SAS 9.1, IBM WS MQ 6.1, WS MB 5.1, Netezza v3.1
- Composants natifs pour les composants de fédération (EII)
- SFTP Enterprise Stage (Secure data transmission)
- Connecteur natif iWay Enterprise:
  - Integration avec 250 sources legacy
- Connecteur SAP R/3 : support SAP ECC6, parallèle (multi-canaux), Appel fonctions RFC, full unicode, support SAP XI
- Les “Rich Connector” (ODBC, Teradata, MQ, ...)





# Évolution de l'interface graphique

- Fusion du Manager dans le Designer:
  - Editeur de fichiers de configuration, gestion des DataSets (EE)
  - Message Handler Manager
  - Import/Export des métadonnées MetaBroker/Bridges
  - Web Service Definition Import
  - Import IMS definitions
  - JCL templates editor (MVS)
  - Import/Export DSX/XML
  - Assistant d'export: 1 composant, tous les composants d'un/des répertoires, objets dépendants...



## ■ Nouvelles fonctionnalités du Designer

### – Arborescence + Référentiel

- Création/gestion libre des répertoires
- Recherche d'objets dans le référentiel (type, date de création, utilisateur, dernière modification...)
- Gestion de plusieurs utilisateurs sur le même job (write/read only)

- Nouveautés : le CDC unifié avec DataMirror
  - Composants dédiés aux problématiques de capture des changements / réplication
- Composants dédiés qualité

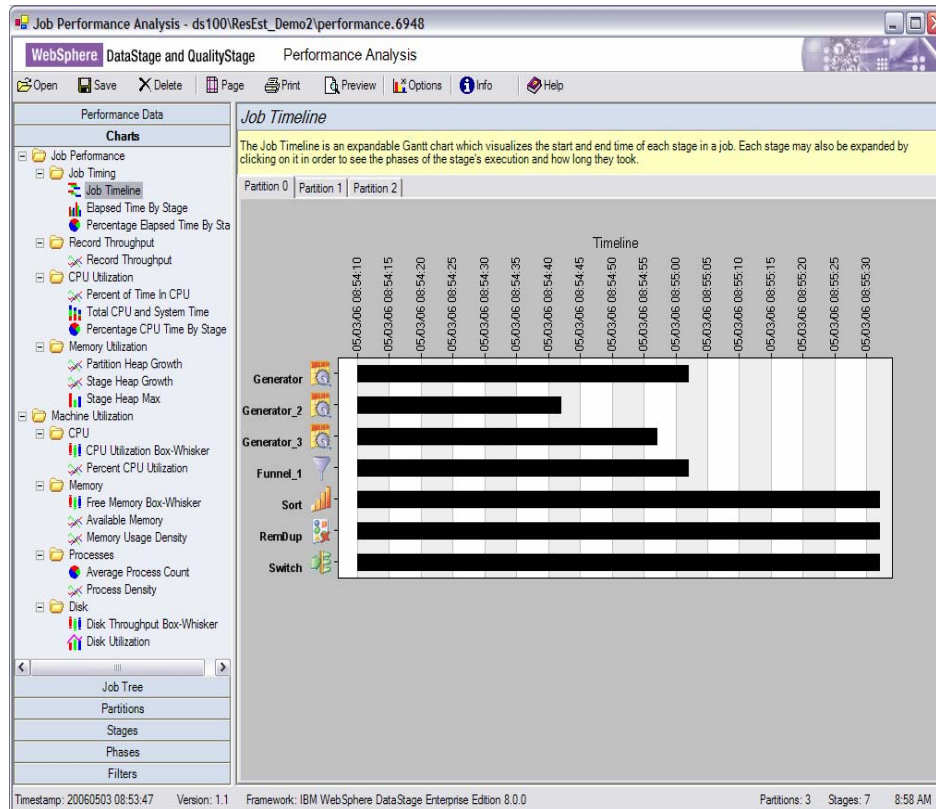
- Investigation, Rapprochement, Standardisation, Dédoublonnage



IBM QualityStage

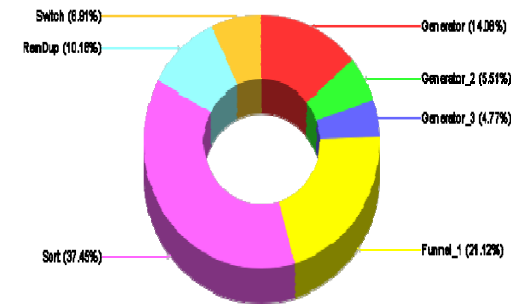
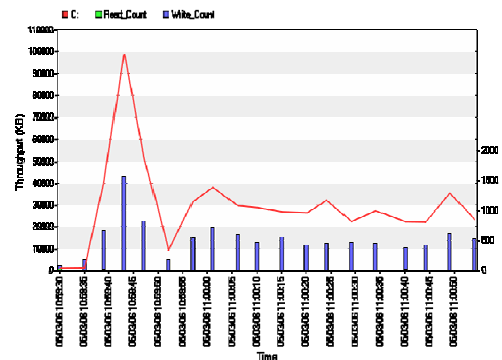
- Nouveautés sécurité
  - Répertoire sécurité: utilisateurs – groupes – rôles
    - Interne : création / mise à jour / suppression
    - Externe : LDAP, Active Directory, Unix
  - Rôles
    - Information Server: Suite User, Suite Administrator
    - Profil par Service : ex: utilisateur DataStage
    - Profil par projets: ex: utilisateur Information Analyzer
  - Authentification basée sur des standards
    - JAAS (Java Authentication and Authorization Service)

# Analyse de performances

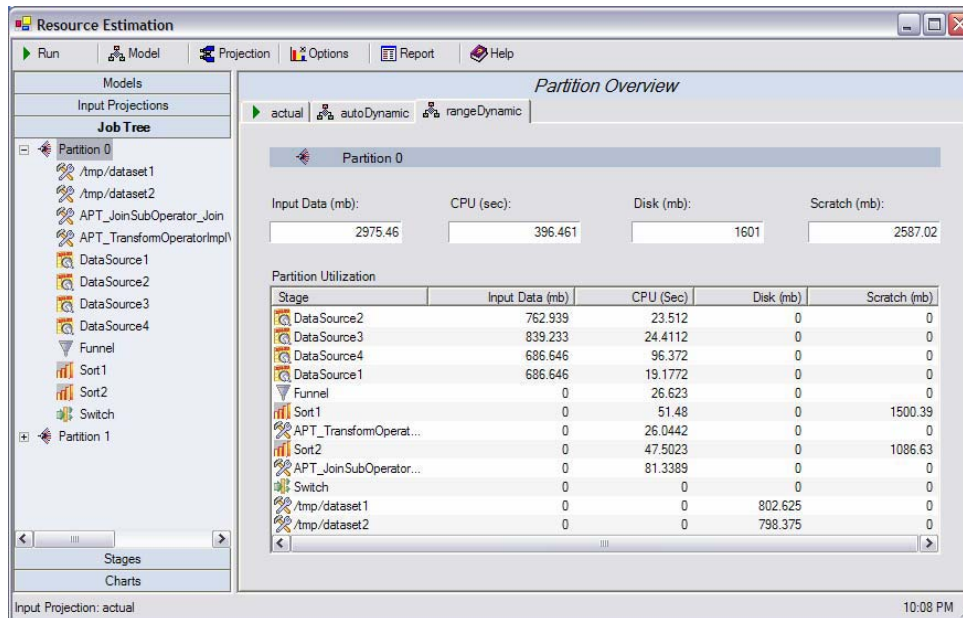


## Un outil dédié et intégré à DataStage :

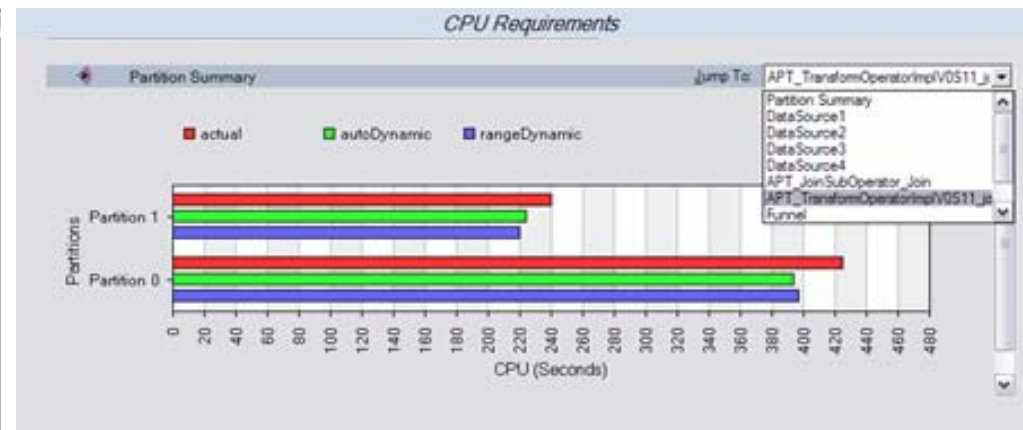
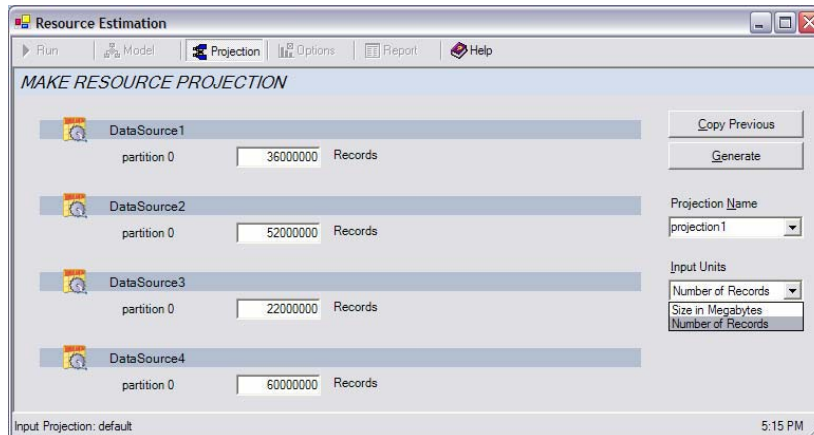
- Pour fournir une vue détaillée des performances et du comportement des jobs DataStage.
- Offre plusieurs modes de visualisation des performances :
  - Par enregistrement
  - Utilisation CPU
  - Temps de réponse du job
  - Mémoire utilisée
  - Mémoire physique utilisée
- Plusieurs graphes pour présenter les résultats



# Estimation des ressources nécessaires



- Donne une **vue synthétique des ressources nécessaires** (CPU, Scratch disk, etc.)
- Permet de **prévoir les ressources** qui seront nécessaires lorsque les volumes de données varient.
- Permet **d'éviter les erreurs de lancement** de jobs en amont lorsque les ressources sont insuffisantes.



# IBM Information Server FastTrack

*Designed specifically for analysts, data modelers and developers to accelerate the translation of business requirements into data integration projects that solve strategic business problems*

Built based on four guiding principles:

1. Streamlining team collaboration
2. Supporting data governance and auditing capabilities
3. Increasing productivity
4. Improving project success and on-time rates

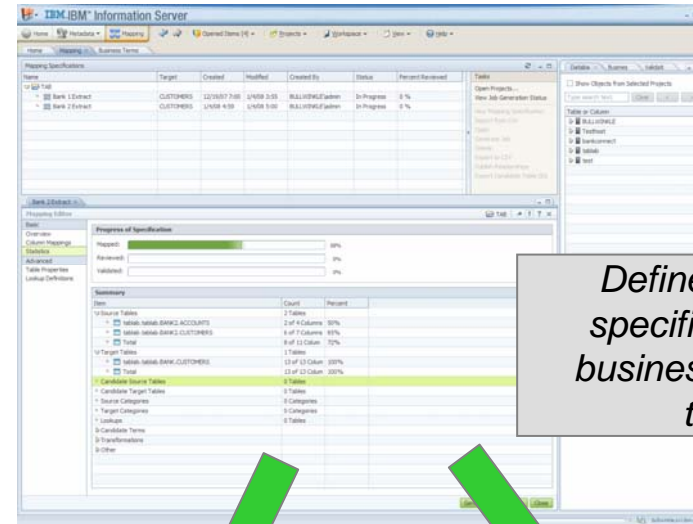


The key to success is in breaking down traditional barriers between manual coordination efforts across multiple user roles and products and putting the emphasis where needed, solving business problems.



# Tracking business requirements to application deployment

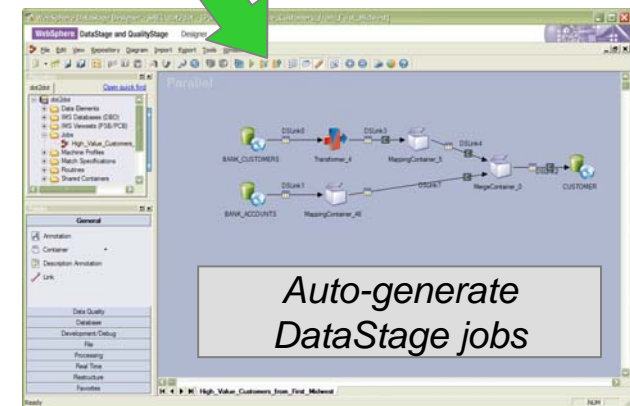
- Single managed infrastructure for data integration projects to track requirements to deployment
- Create required transformation rules for source-to-target mappings
- Define and link business terms to physical structures
- Generate DataStage jobs
- Generate historical documentation for tracking



*Define mapping specification with business rules and terms*



*Flexible reporting and tracking*



*Auto-generate DataStage jobs*

# Maximizing Business Value

Customers identified top 3 FastTrack impacts on organizational processes:

## Faster Time To Value

- **Common format** for requirements documentation
- **Remove ambiguity** in project specifications
- Track **historical specifications** for auditing
- **Improve communication** across team members

## Increased Productivity

- **Jump start ETL** by automatically translating business logic into code
- **Synchronize work** between disparate development teams
- Improve success of **off-shore ETL development models**
- **Share common metadata** between analyst and developer without extra work

## Stronger Governance

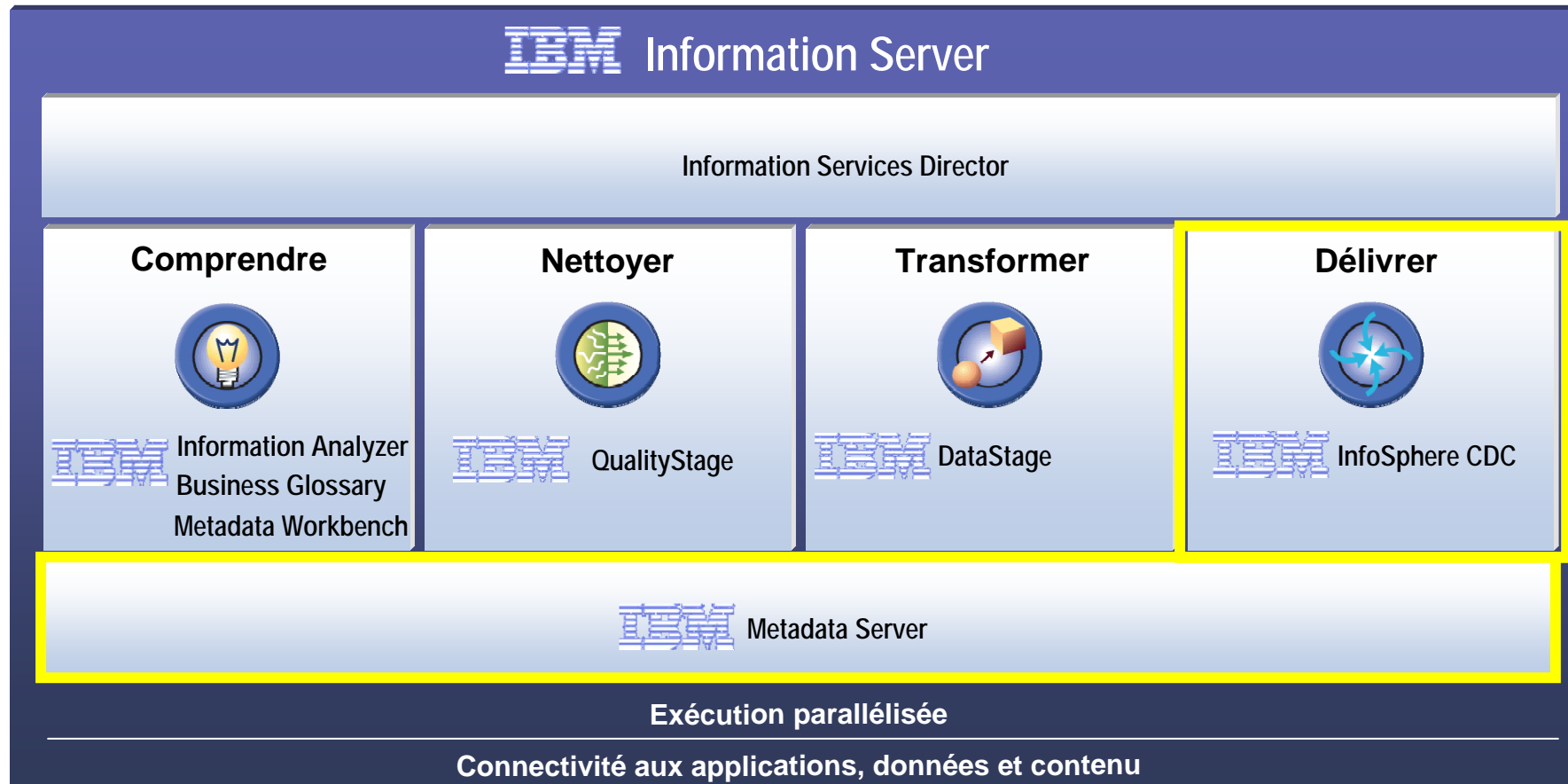
- **Centrally managed** location to create, store, maintain and audit start to finish
- Ensure **standards and documentation** control are reinforced
- Expose critical links between **transformation rules and business requirements**
- **Eliminate storage of Excel files** and hidden rules

*“FastTrack brings significant productivity gains, quality improvement and metadata transparency to integration environments – up to **40%** reduction in time.”*

*– Beta Customer*

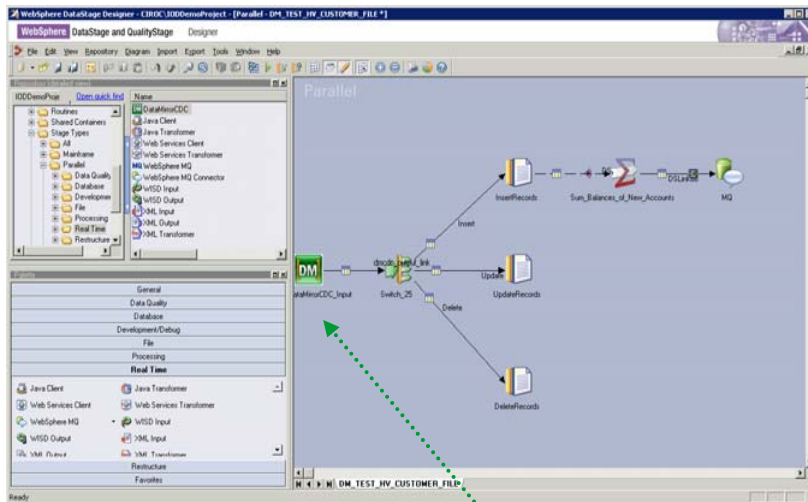
# IBM Information Server

*Delivering information you can trust*



# IBM information Server Change Data Capture

Information Server Change Data Capture

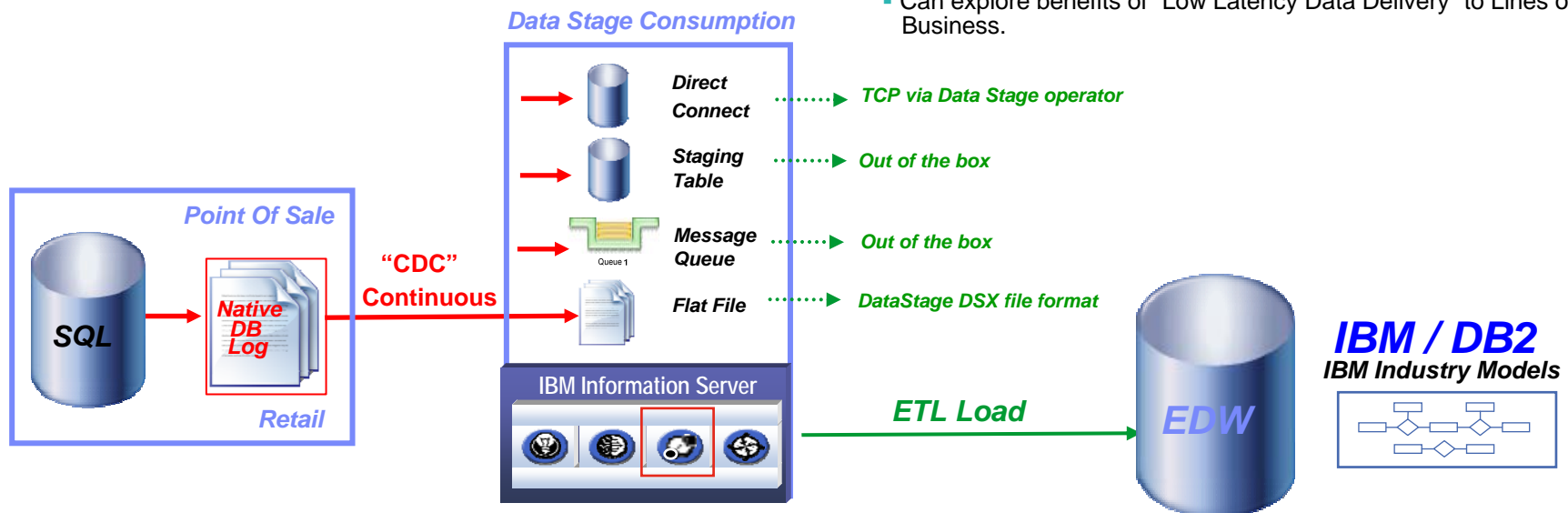


## Technical Benefits

- Extended breadth of native CDC source coverage.
  - Now consider CDC for SQL, Oracle, i5/DB2, Sybase.
- Defined integration roadmap (First delivered Q1 2008).
  - Integrated with DataStage Designer.
  - Metadata integration for lineage, impact analysis.
- Flexibility - Four options to stream OLTP changes to Datstage.

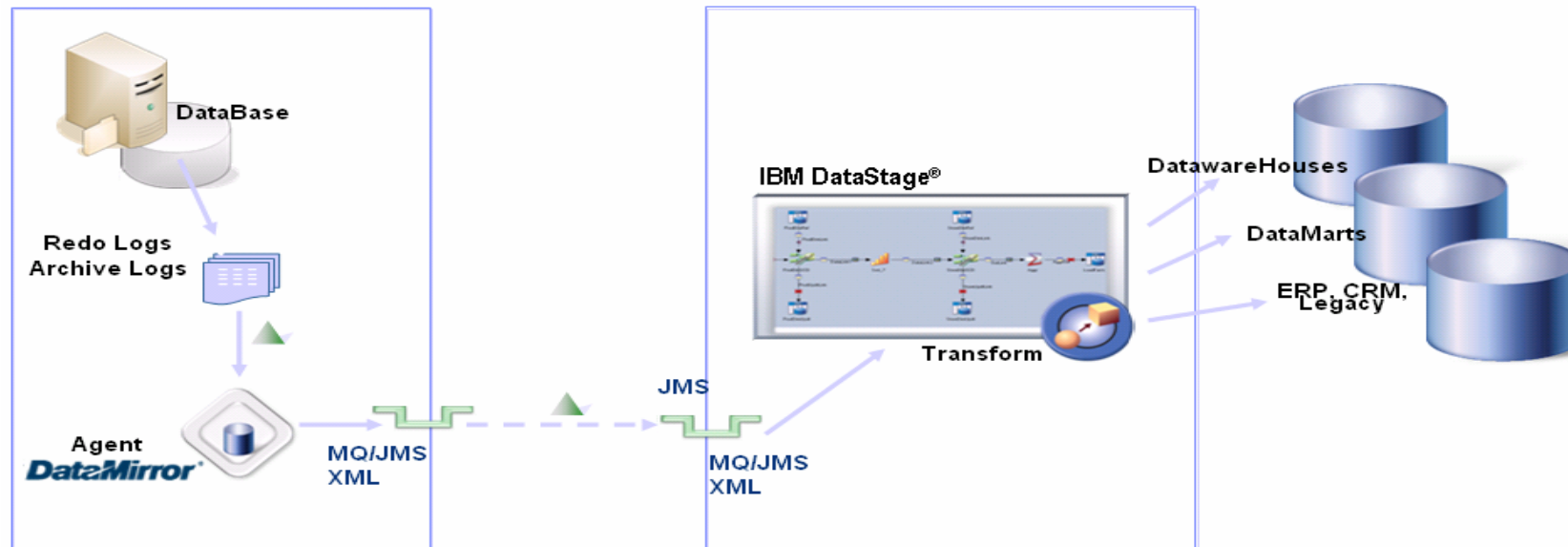
## Why Should Companies Care?

- Reduces risk on production systems (low impact).
- Allows IT to better manage batch windows, network loads.
- Can explore benefits of "Continuous ETL" (i.e. Active EDW)
- Can explore benefits of "Low Latency Data Delivery" to Lines of Business.



# IBM information Server Change Data Capture

## Architecture Temps Réel



**Databases**

Oracle, DB2, DB2 UDB, SQL Server, Sybase, Teradata, Netezza, PointBase

**Platforms**

z/OS, System i5, Red Hat and SUSE Linux, AIX, HP/UX (PA-RISC and Itanium), Solaris SPARC, Tru64 UNIX, Windows

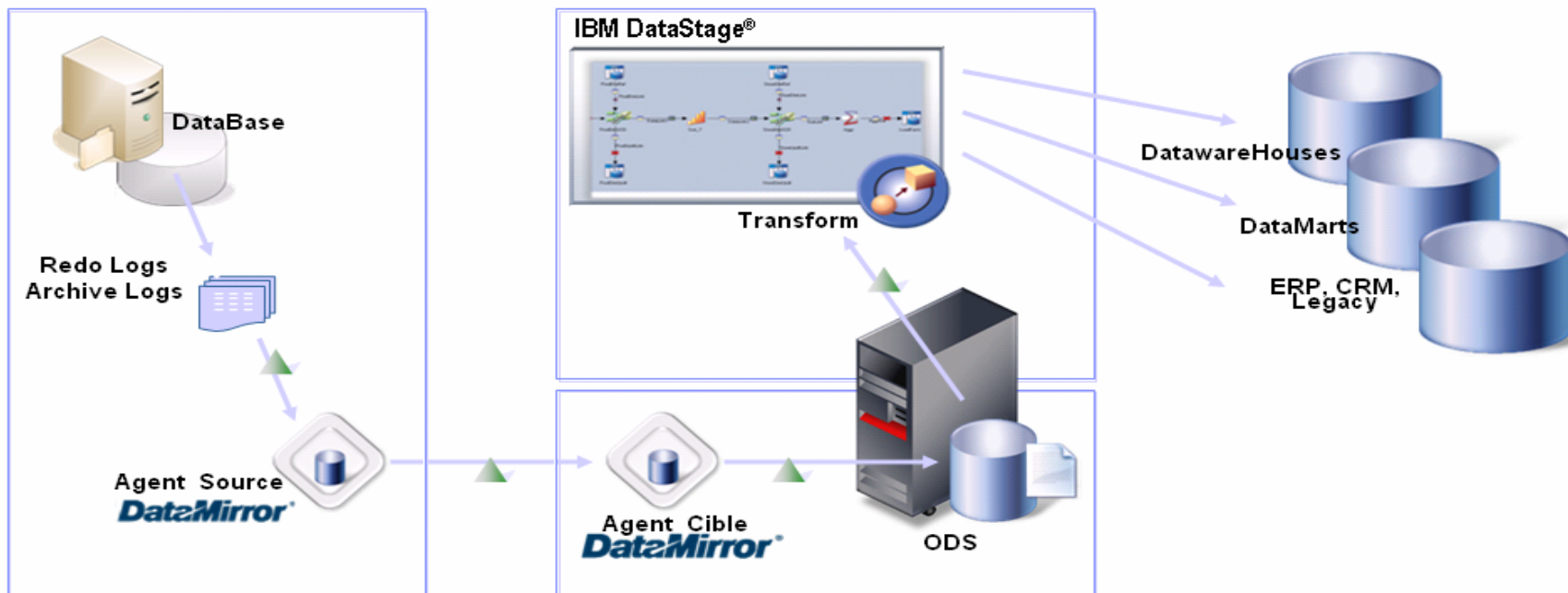
**Messaging Middleware**

MQSeries, Sun Open Message Queue (JMS), TIBCO, BEA AquaLogic, Oracle Fusion Middleware



# IBM information Server Change Data Capture

## Architecture Batch



### Databases

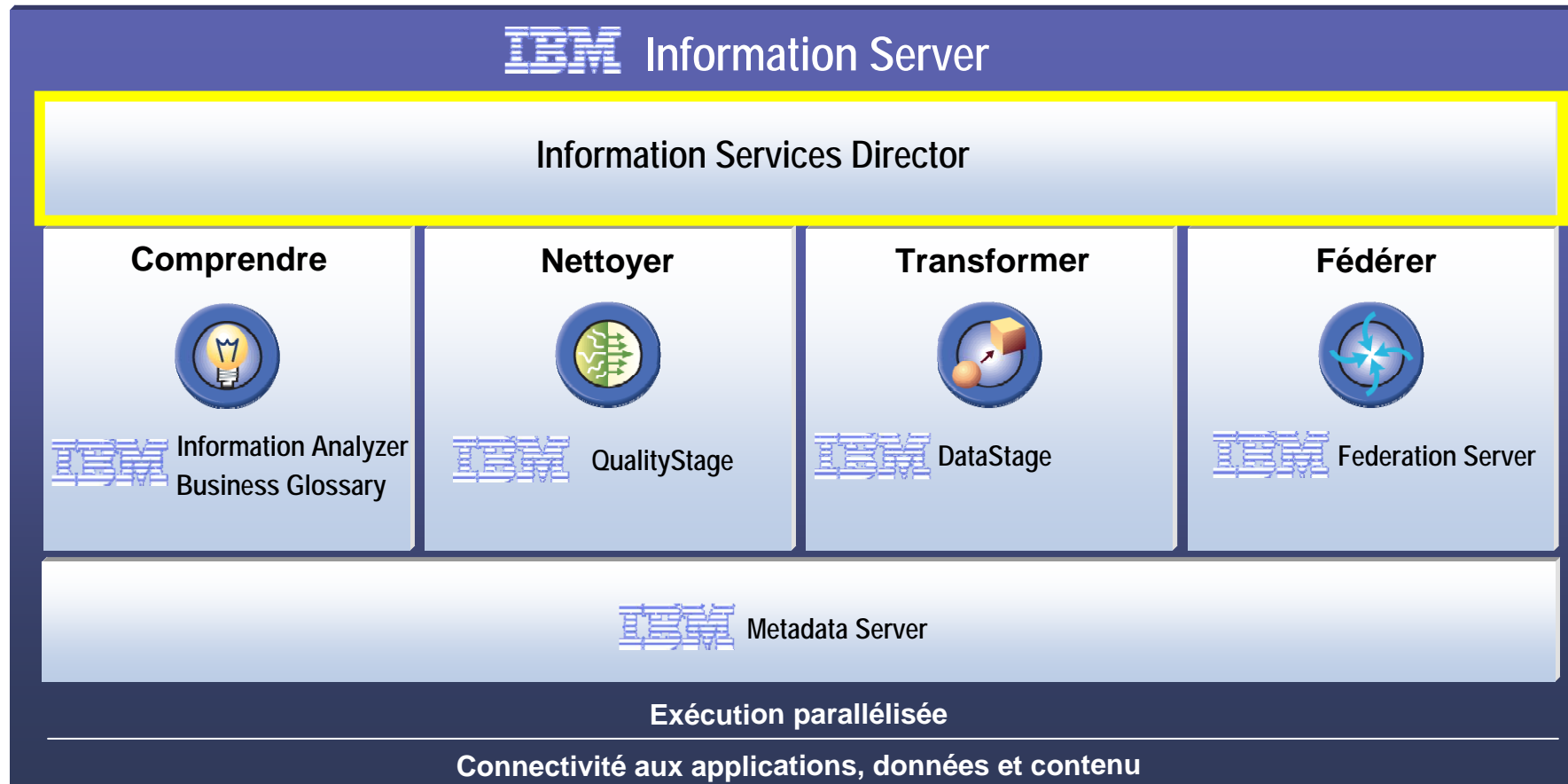
Oracle, DB2, DB2 UDB, SQL Server, Sybase, Teradata, Netezza, PointBase

### Platforms

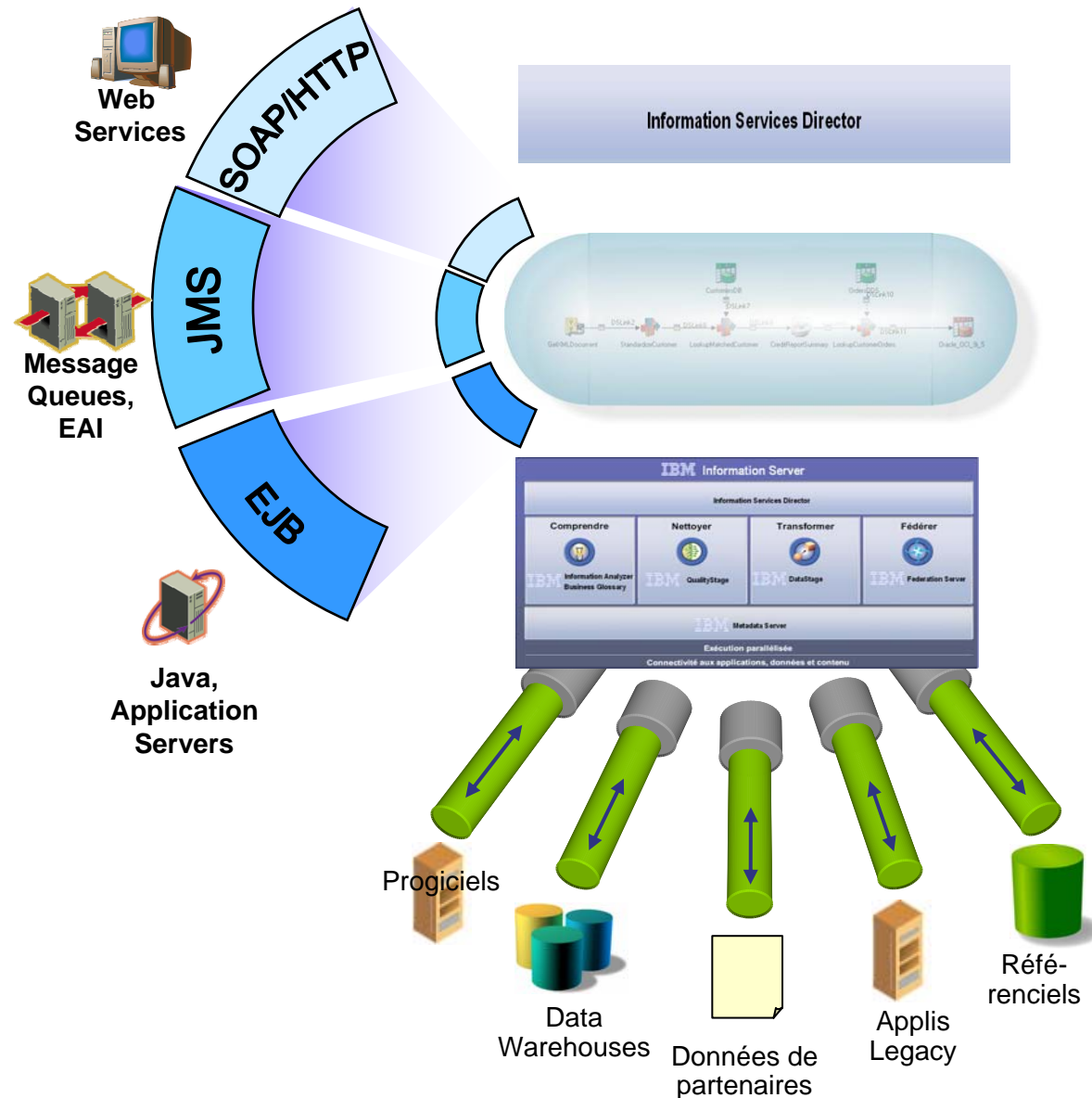
z/OS, System i5, Red Hat and SUSE Linux, AIX, HP/UX (PA-RISC and Itanium), Solaris SPARC, Tru64 UNIX, Windows

# IBM Information Server

*Delivering information you can trust*



# Déploiement de Services ETL à la demande



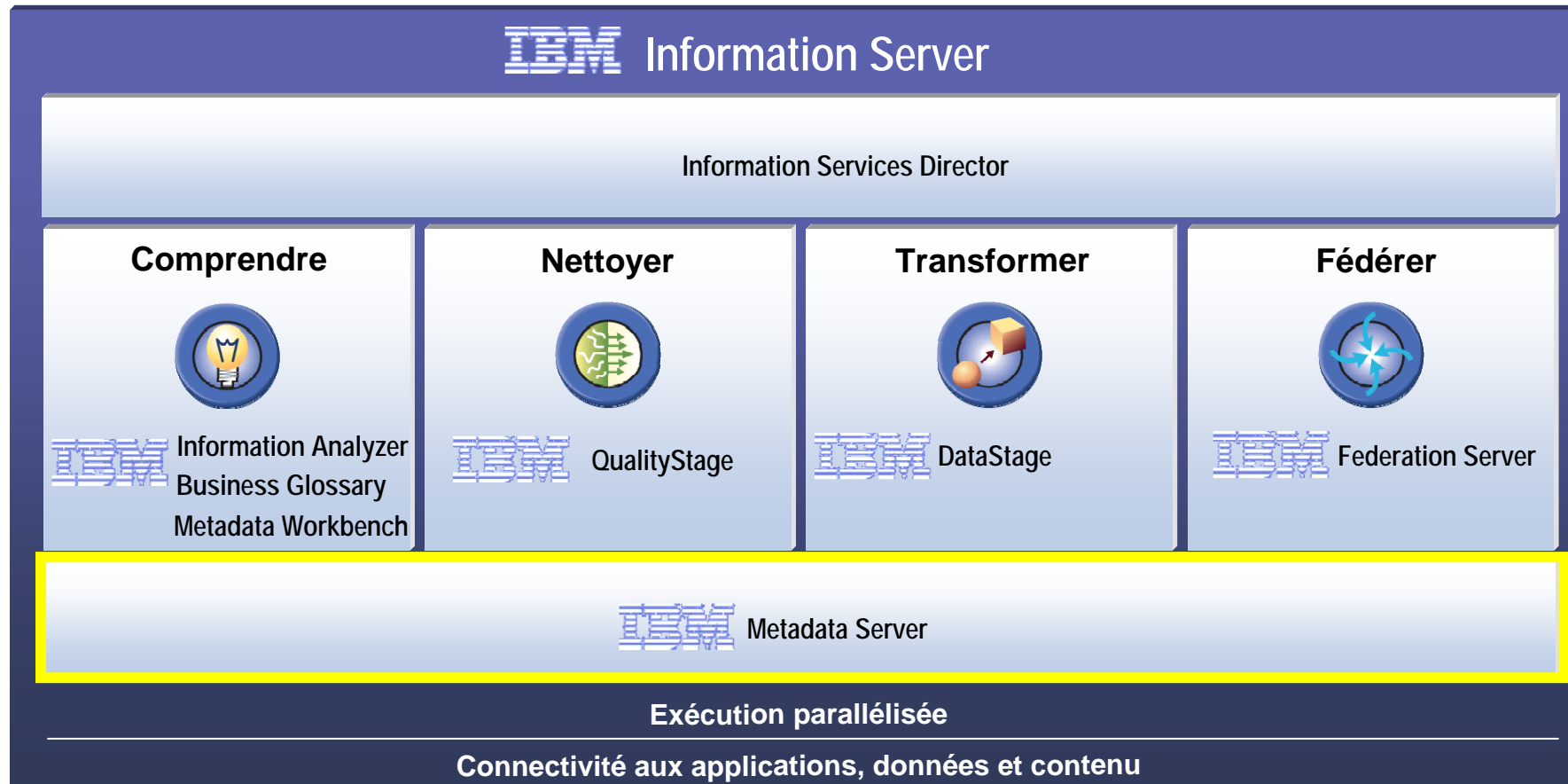
- « Package » la logique d'intégration de donnée (jobs ou maps) en services qui isolent les développeurs des technologies employées
- Permet à ces services d'être invoqués via EJB, JMS, ou Web Services (SOAP/HTTP)
- Apporte une fonction d'équilibrage de charge et de distribution du traitement sur un ou plusieurs serveurs DataStage
- Intègre les fondations d'une infrastructure de services (sécurité, logging, monitoring)

# Information Server: Enterprise Service Bus integration



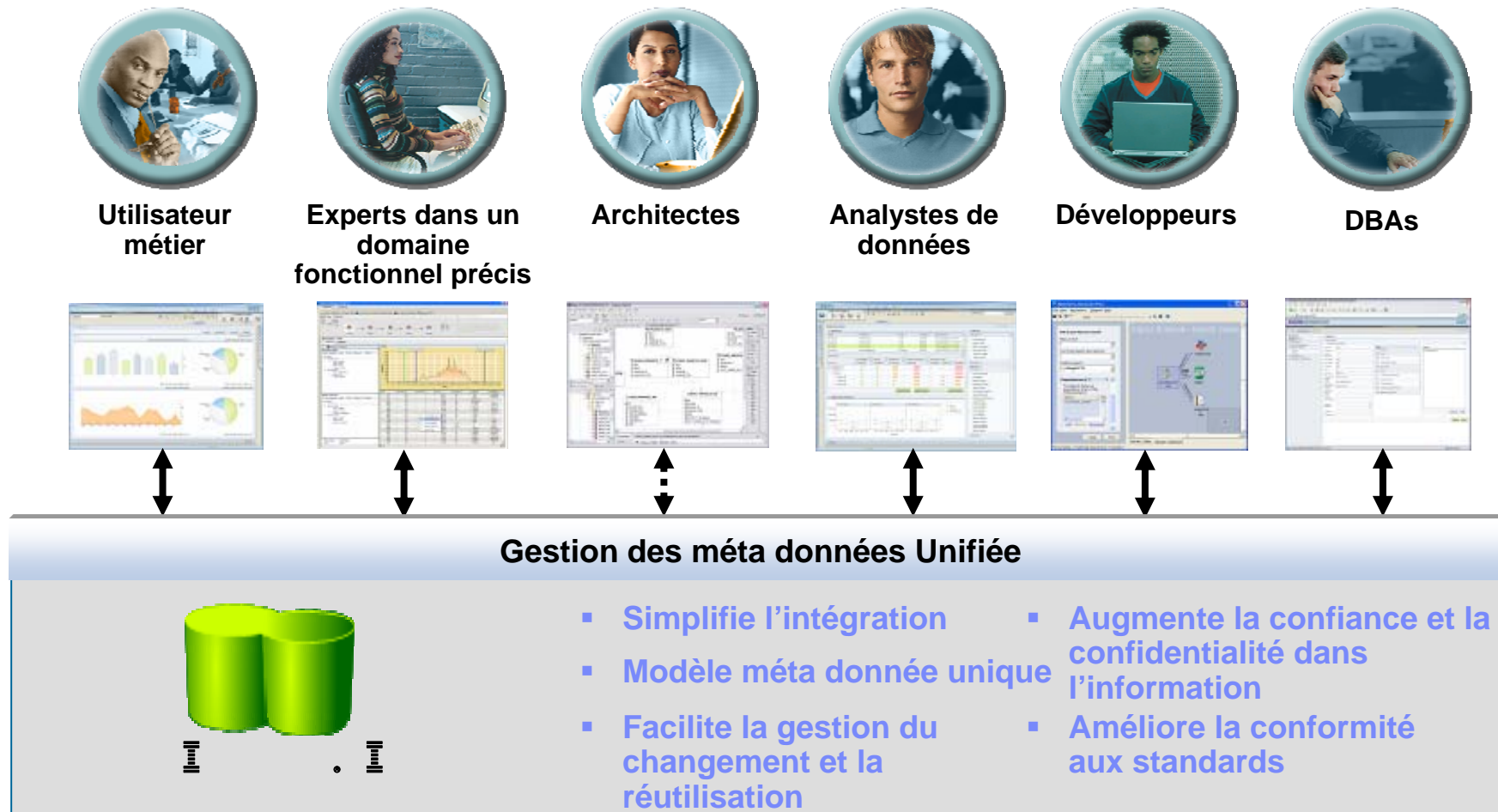
# IBM Information Server

*Delivering information you can trust*





# Plateforme orientée Rôle avec gestion des méta données intégrée



# Méta données métier : Business Glossary

- Application Web permettant de gérer et partager les méta données métier
- Alignement des efforts IT vers des objectifs métier
- Fournir un contexte métier à des capital technique du SI
- Etablir une responsabilité de l'information
- Passerelle entre les utilisateurs métiers et les interlocuteurs techniques

Base = DB2

Schema = NAACCT

Table = DLYTRANS

Colonne = ACCT\_NO

Type = char(11)



Technical



Business

## Numéro de compte

Les 10 digits du numéro de compte.

Quelques fois se réfèrent à un ID de compte.

Sa valeur a le format L-FIIIVVVV.



**Comprendre**



**Experts dans un domaine fonctionnel précis**



**Utilisateur métier**

**WebSphere Business Glossary**

**Créer et gère le vocabulaire métier, et le lie au méta données techniques**



**Vue métier**

# IBM Business Glossary : Objectifs

- Un Outil web pour les utilisateurs Métiers permettant de :
  - Créer et gérer un vocabulaire commun et contrôlé
  - Créer & gérer des règles hiérarchiques entre les données (Business Taxonomy)

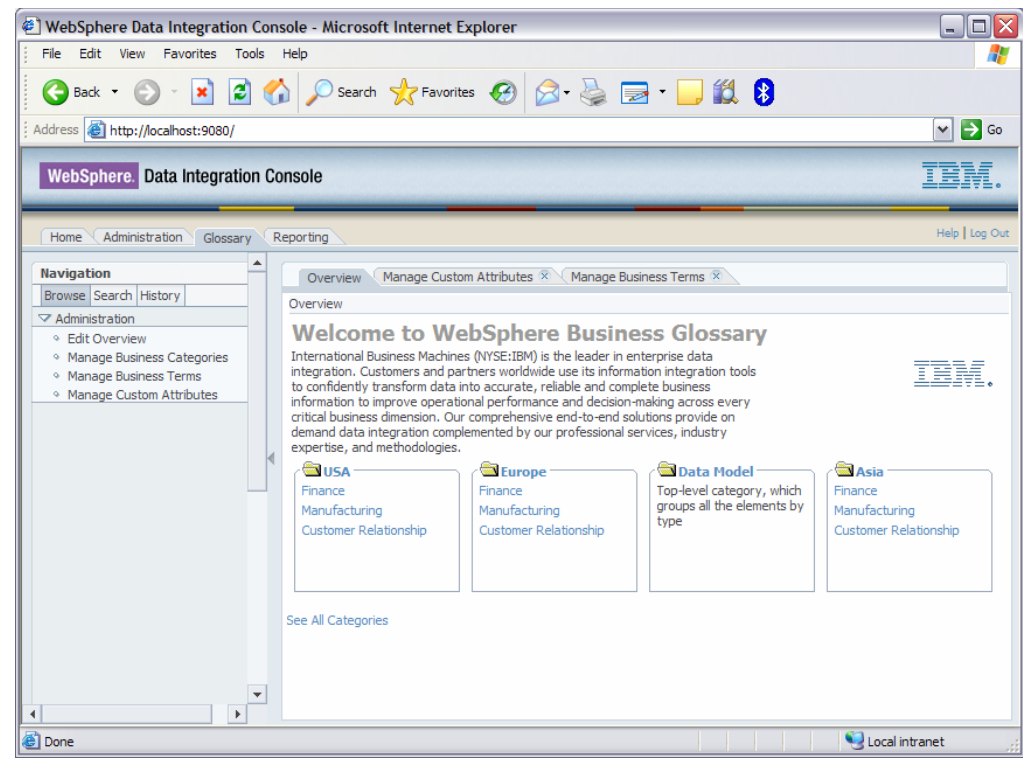
```

Finance
- United States
- General Ledger
- Account Number
- Account Type
    
```

- Gestion Collaborative des méta données métier

- Référentiel pour tous les assets de l'entreprise

- Signification
- Description
- Contexte d'utilisation
- Synonyme
- Notes
- Editeur/Gérant




## Qu'est-ce que IBM Metadata Workbench?

- Exploration des méta données générées et utilisées au sein d'Informations Server
- Compte rendu sur les données, leurs mouvements, leurs significations métiers, leurs dépendances et leurs impacts, leurs locations physiques, etc...
- Tracer les données provenant des rapports métiers à travers des applications éparpillées (Process, Quality Center, ETL, Data Marts, etc....)



**Manager  
d'intégration**



**Developpeurs**

**Comprendre**



**IBM Metadata Workbench®**

Fournit aux professionnels IT un outil pour explorer et comprendre les informations produites et employées par la suite Information Server.



# Fonctionnalités de Metadata Workbench

## EXPLORATION

**Explore** les informations clés :

Jobs, Rapports, Bases de données, Modèles, Terminologie, questionnaires, Systèmes

- ✓ Navigation simple à travers les informations clés
- ✓ Recherche simple & avancée (multi-critères)
- ✓ Vue consolidée de l'information utilisée dans toute la suite.
- ✓ Vue graphique des relations.
- ✓ Requeteur Ad-hoc (query builder)

## ANALYSE

**Analyse** Les dépendances et les relations entre les informations clés.

- ✓ Trace les mouvements de données provenant et allant vers les bases de données, les jobs et les rapports.
- ✓ Aide à la compréhension de la signification des colonnes, tables et autres informations.
- ✓ Aide à la compréhension de l'impact sur les changements sur une colonne ou d'un job à travers Information Server.

## GESTION

**Gère** les méta données d'Information Server pour permettre une analyse approfondie.

- ✓ Edition des noms et description de chaque méta donnée.
- ✓ Création de gestionnaire de méta données.
- ✓ Réconciliation des informations dupliquées.
- ✓ Accès aux méta données de runtime pour enrichir le reporting.



WebSphere Metadata Workbench - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Refresh Home Search Favorites

Address http://212.179.34.142/unicorn/index.js Go

IBM. Metadata Workbench

Home Workbench Help About Log Out

### Report Finder - Data Movement

Data Movement for: Customer Name Attribute

```
graph LR; C1[Column Definition CUSTNAME1] -- Used As Flow Variable --> J1[Job Flow Variable CUSTNAME1]; C2[Column Definition CUSTNAME2] -- Used As Flow Variable --> J2[Job Flow Variable CUSTNAME2]; J1 -- Is Source Of Flow Variable-Created From Data Field --> CD[COLUMN DEFINITION CUST_NAME]; J2 -- Is Source Of Flow Variable-Created From Data Field --> CD; CD -- Used As Flow Variable --> J3[Job Flow Variable CUST_NAME]; J3 -- Is Source Of Flow Variable-Created From Data Field --> DF[Data Field DWCUSTNAME]; DF -- Report Attribute --> RA[BI Report Attribute Customer Name Attribute];
```

### Report Finder - Business Meaning Report

Business Meaning Report for: ODS\_ORD

```
graph LR; BT[Business Term Order] -- Referenced by other Object --> DME[Data Model Entity Orders]; DME -- Implemented By Data Collection --> DC[Data Collection ODS_ORD];
```

### Report Finder - Asset Dependency Report

Asset Dependency Report for: DataWarehouse

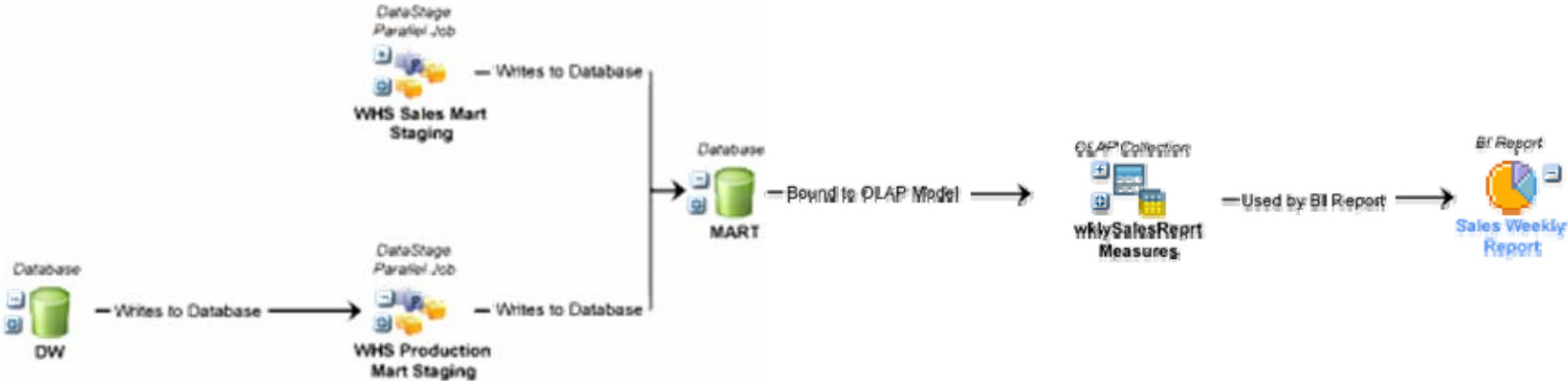
```
graph LR; HS[Host System DataWarehouse] -- Hosts Data Store --> DB[(Database DW)]; DB -- Has Data Schema --> DS[Data Schema DW];
```

•Tracabilité sur les objets Métiers.

•Tracabilité au niveau physique

Internet

# IBM InfoSphere Information Server & Cognos 8 BI: End-to-End Data Lineage



Data Warehouse and Source Systems
DataStage & QualityStage Job
Data Mart
OLAP Collection
Report

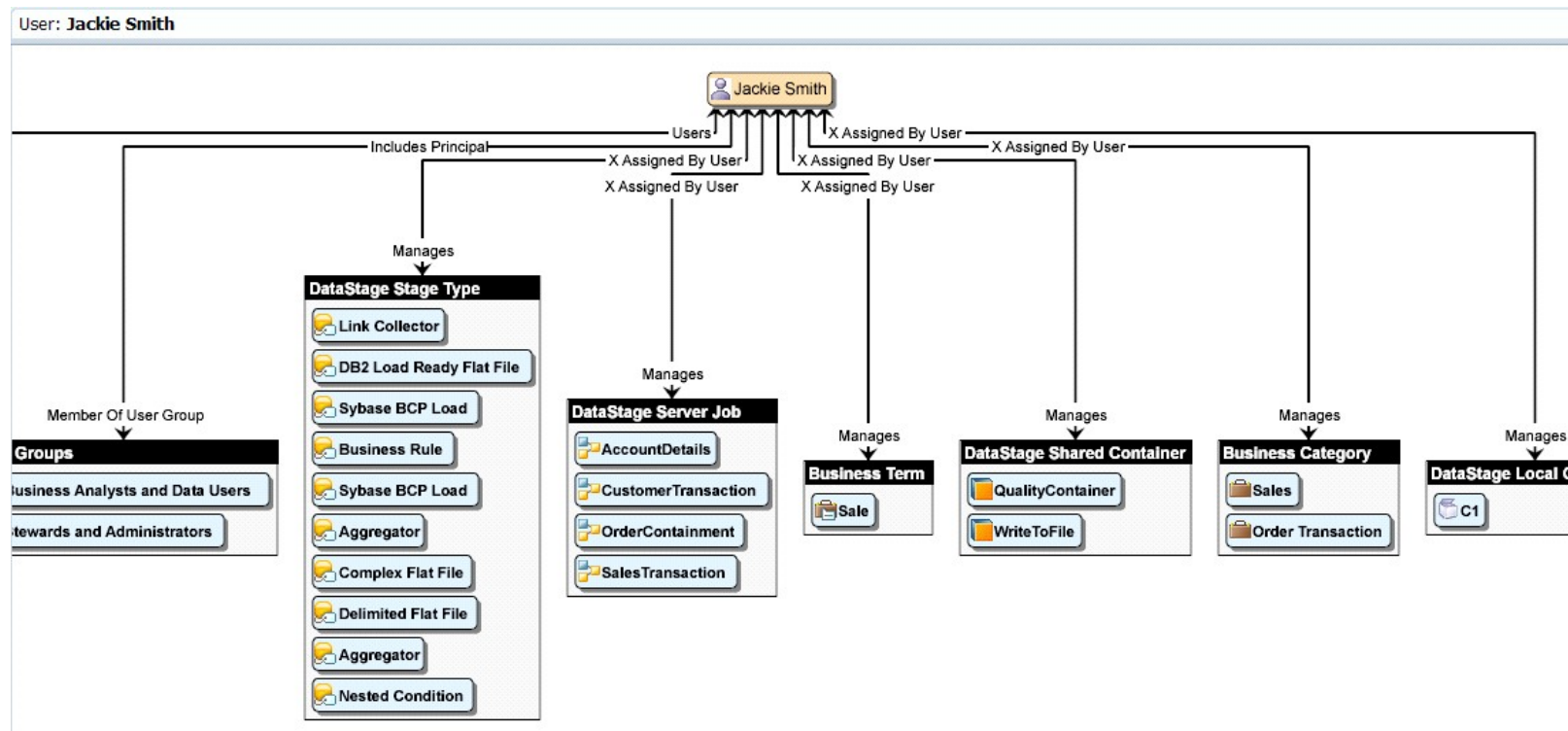


**IBM.** Information Server

**COGNOS**  
AN IBM COMPANY

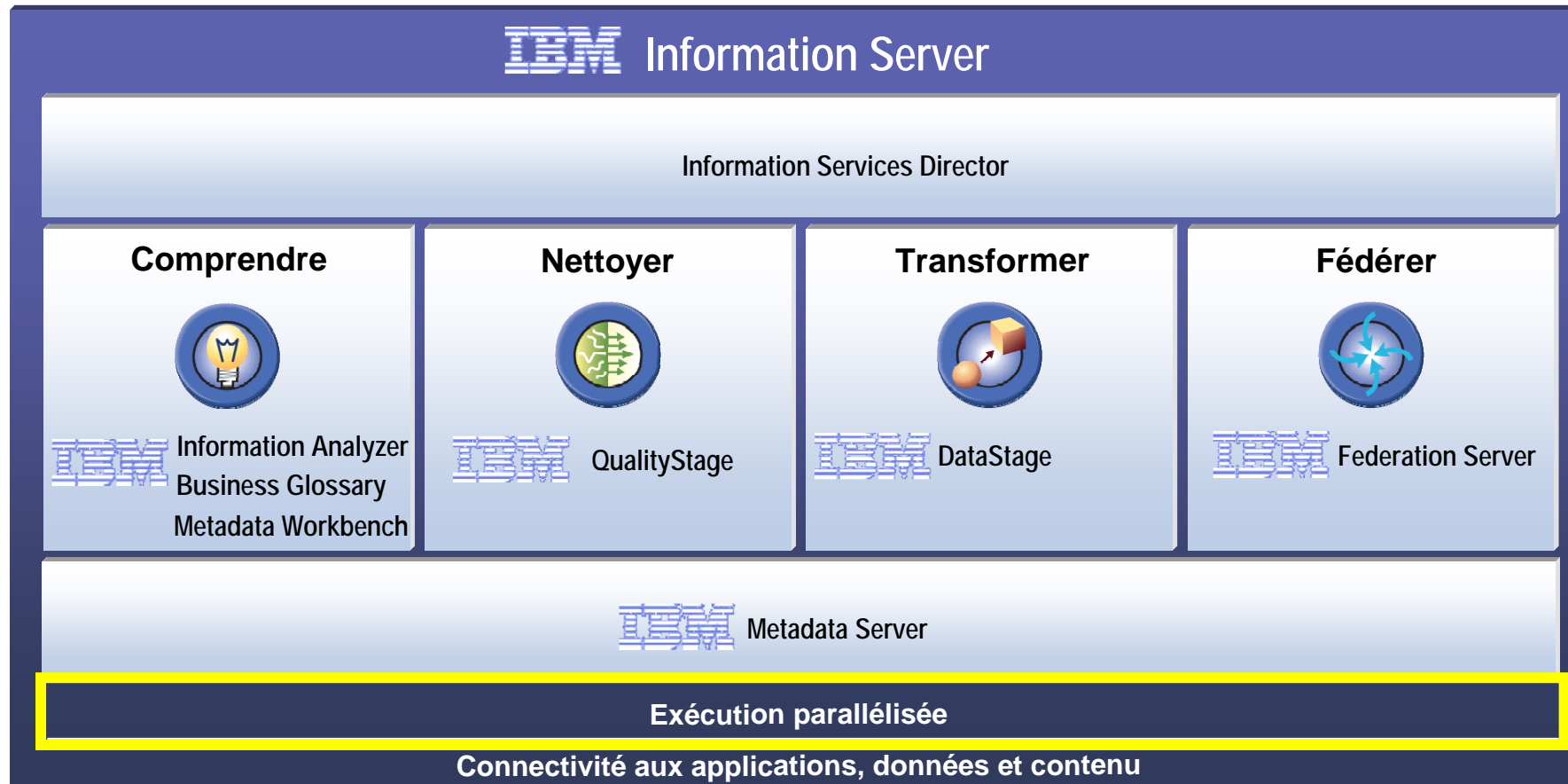
# Meta Data Workbench

- Responsabilité/Gérant



# IBM Information Server

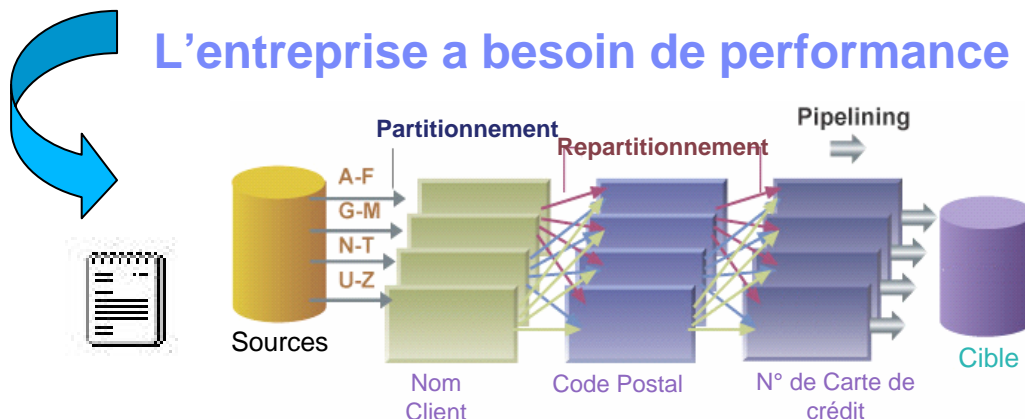
*Delivering information you can trust*



# Performances: Exécution Parallélisée



L'entreprise a besoin de performance



Un fichier de configuration décrit la topologie d'exécution.

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

### Le Pipelining

- Mode rapide (mémoire) d'acheminement des Données.

### Le Partitionnement

- Une autoroute multivoies, bout en bout.

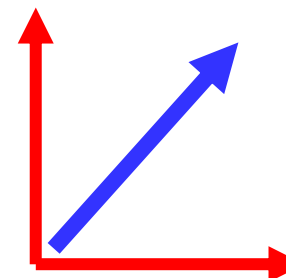
### Le Repartitionnement

- Capacité à changer de voie, à la volée, en fonction des actions à réaliser

## A Propos de Enterprise Edition

- Réalisation des processus d'intégration sans nécessité de prendre en compte les volumes de données et contraintes de temps
- Etapes simplifiées permettant de définir des partitionnement multiples dynamiques dans chaque processus si nécessaire
- Simple fichier de configuration pouvant être modifié pour ajouter de nouveaux processeurs
- Pas de programmation manuelle pour prendre en charge d'avantage de processeurs
- Support et environnements SMP, Cluster, GRID, et plates-formes MPP

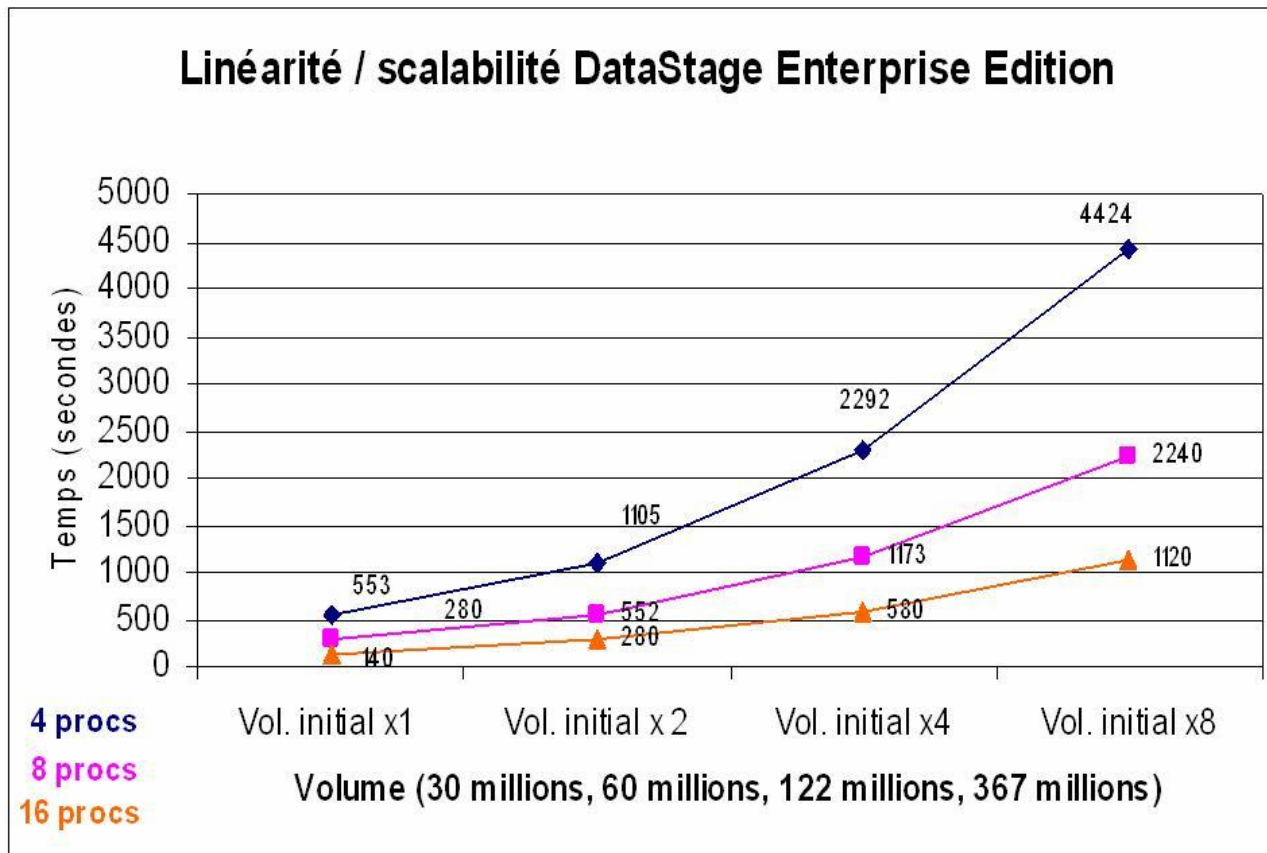
## Linéarité des Performances





# Linéarité des performances de Datastage

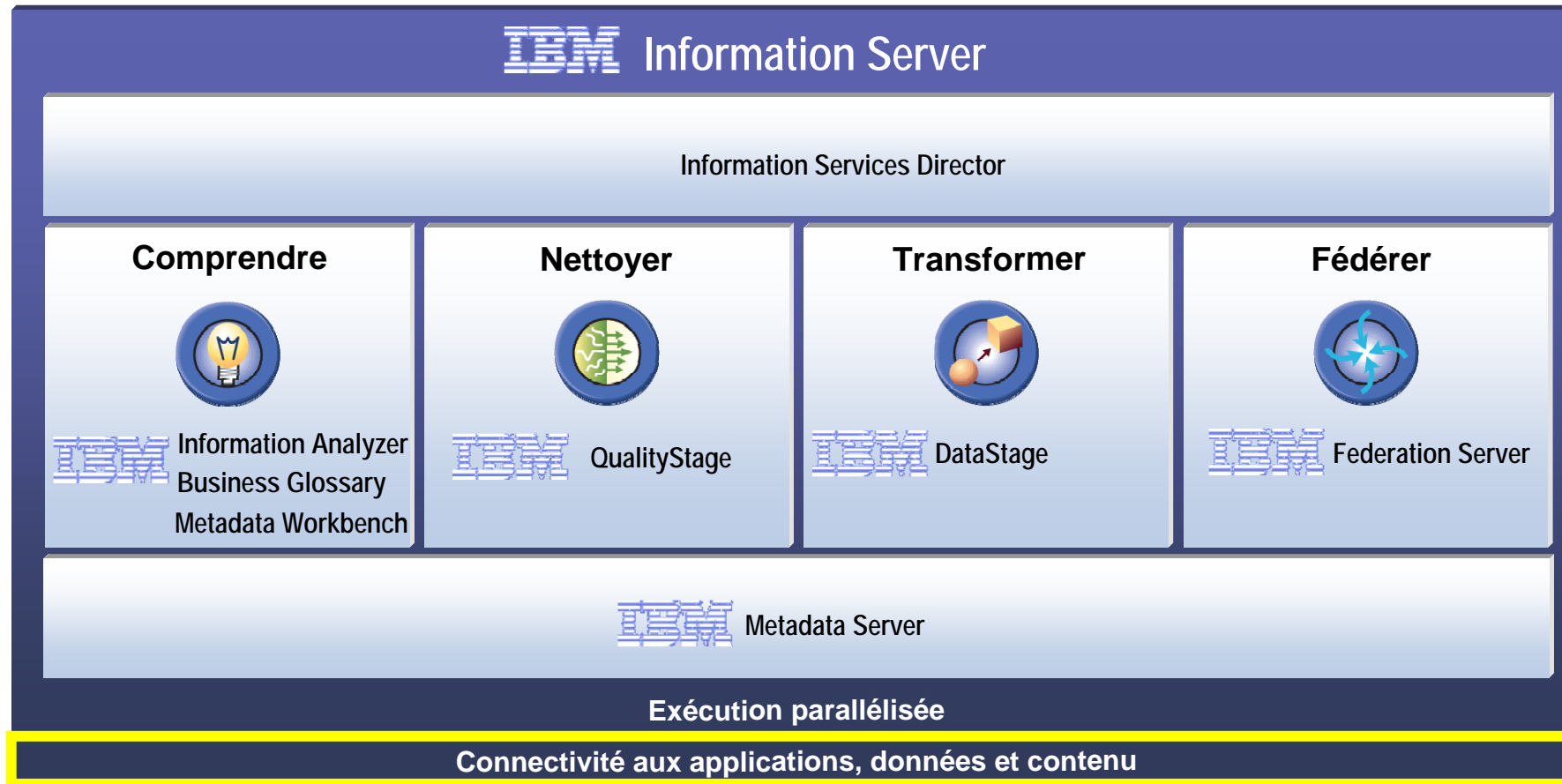
Transformer



- **Linéarité totale:**  
Volumétrie - Nombre de CPUs
- **Parallélisation**  
répartie automatiquement selon les noeuds
- **Scalabilité:**  
Prise en compte auto. de modification de config. matérielle (ajout de cpus, disques) par fichier de paramétrage

# IBM Information Server

*Delivering information you can trust*



# L'importance de la connectivité

## RDBMS

DB2 (on Z, I, P or X series)  
 Oracle  
 Informix (IDS and XPS)  
 Ingres  
 MySQL  
 Netezza  
 Progress  
 RDB  
 RedBrick  
 SQL/DS  
 SQL Server  
 Sybase (ASE & IQ)  
 Teradata  
 Universe  
 UniData  
 NonStopSQL  
 And more.....



## General Access

Sequential File  
 Complex Flat File  
 File / Data Sets  
 Named Pipe  
 FTP  
 Compressed / Encoded Data  
 External Command Call  
 Parallel/wrapped 3<sup>rd</sup> party apps  
 EMC InfoMover  
 Web logs  
 Unstructured: e-mail, docs, etc.  
 Content Management Systems  
 Life Sciences



## Standards & Real Time

WebSphere MQ  
 Java Messaging Services (JMS)  
 Java  
 XML & XSL-T  
 EBXML  
 Web Services (SOAP)  
 Enterprise Java Beans (EJB)  
 EDI  
 FIX  
 SWIFT  
 HIPAA



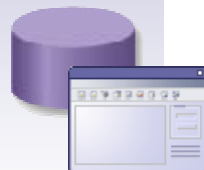
## Legacy

Allbase/SQL  
 C-ISAM  
 D-ISAM  
 Datacom/DB  
 DS Mumps  
 Enscribe  
 Essbase  
 FOCUS  
 IDMS/SQL  
 ImageSQL  
 Infoman  
 KSAM  
 M204  
 MS Analysis  
 Nomad  
 Nucleus  
 RMS S2000  
 Supra  
 TOTAL  
 Turbolmage  
 Unify  
 And many more....



## Enterprise Applications

JDE/PeopleSoft EnterpriseOne  
 Oracle Applications  
 PeopleSoft Enterprise  
 SAS  
 SAP R/3 & BI  
 SAP XI  
 Siebel  
 JDA  
 Ariba  
 Manugistics  
 I2  
 And more...

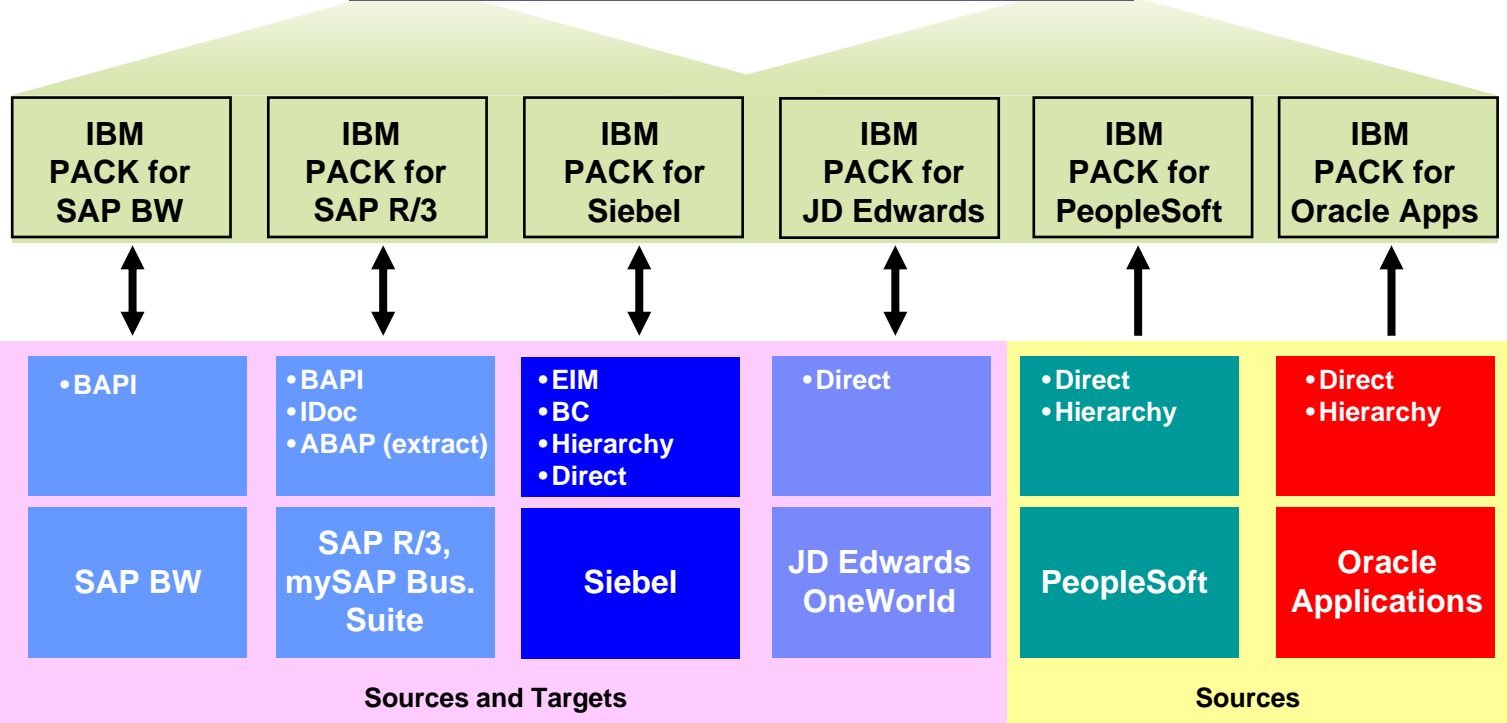
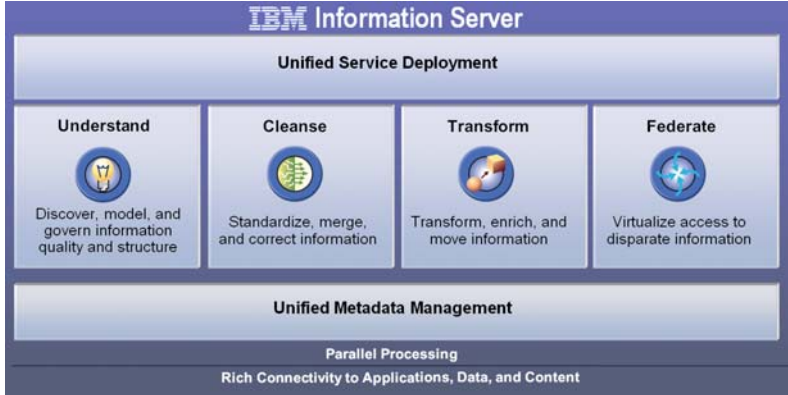


## CDC / Replication

DB2 (on Z, I, P, X series)  
 Oracle  
 SQL Server  
 Sybase  
 Informix  
 IMS  
 VSAM  
 ADABAS  
 IDMS  
 NonStopSQL  
 Enscribe

# IBM Packaged Application Connectivity Kits – PACKs

*Developped by IBM and Certified By SAP.*



## Liste des passerelles d'échanges

### ***Outils de Modélisation***

- ERwin
- PowerDesigner
- ER/Studio
- Oracle Designer
- Popkin
- SilverRun (via UML)
- **Rational Rose (via UML)**
- **Mega (via XML/XMI)**
- Visio
- Visual Studio Modeler

### ***Packages Fichiers, SGBDs***

- ODBC
- Excel
- XML / Schema
- Cobol CopyBook
- Teradata MDS

### ***BI et OLAP***

- Cognos Impromptu
- Cognos PowerPlay
- **Cognos ReportNet**
- Business Objects
- IBM DB2 Cube Views
- MicroStrategy
- Impromptu
- **Hyperion EIS**
- **Hyperion Application Builder**
- **MS Analysis Services**
- SAS Information Map
- SAP BW
- PeopleSoft EPM

### ***Packages UML et CWM***

- OMG CWM / XML
- OMG UML / XMI

### ***ETL***

- **Humingbird Genio**
- BO Data Integrator
- **Informatica**
- MS Integration Services
- IBM Warehouse Manager
- SAS ETL Studio
- Oracle Warehouse Builder



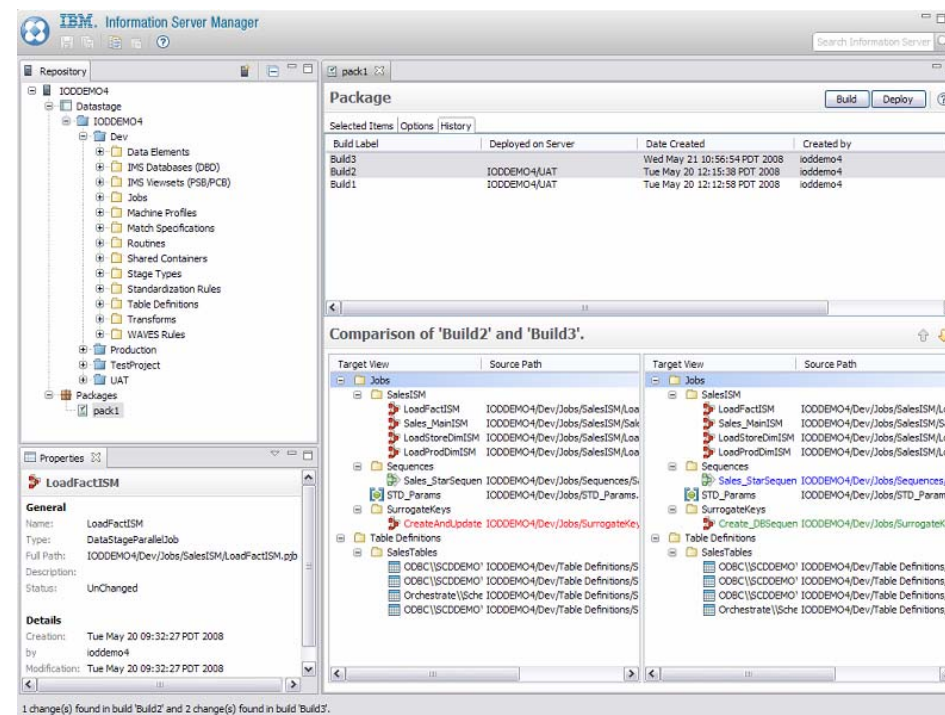
# InfoSphere Information Server Manager Enhanced Deployment

## Compare

- Users can compare deployed builds side by side
- Allows users to see new objects, changes and if any new versions that may be in a subsequent build.

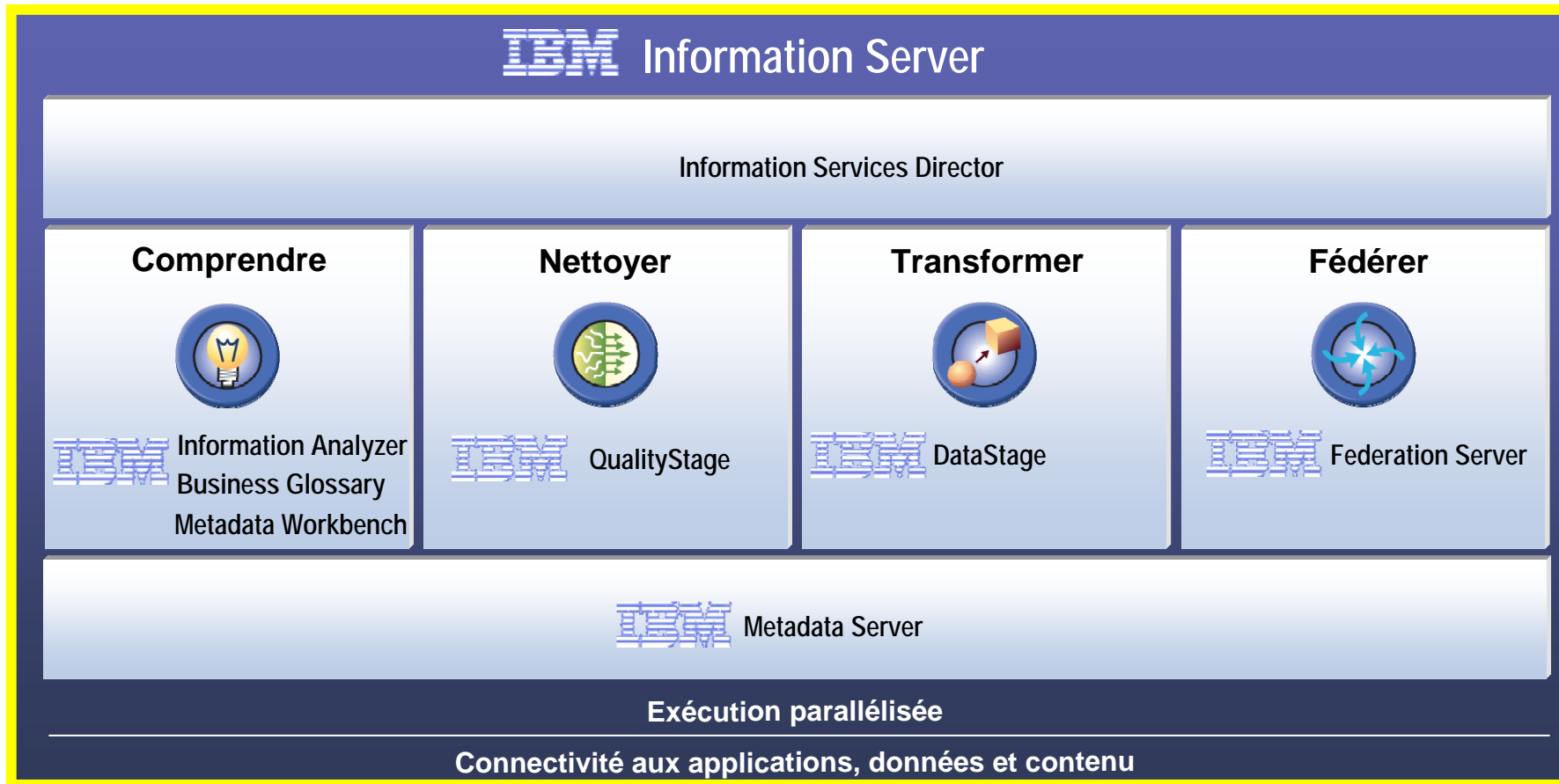
## Audit Trail

- A record is maintained of all the deployments made
  - Who performed the action
  - When it was done
  - and all the objects that were moved in that deployment
- This provides users with the ability to track any movement into production systems



# IBM Information Server

*Delivering information you can trust*



## Conclusion ... Pourquoi passer en V8 ?

### Bénéficier des nouvelles fonctionnalités DataStage v8

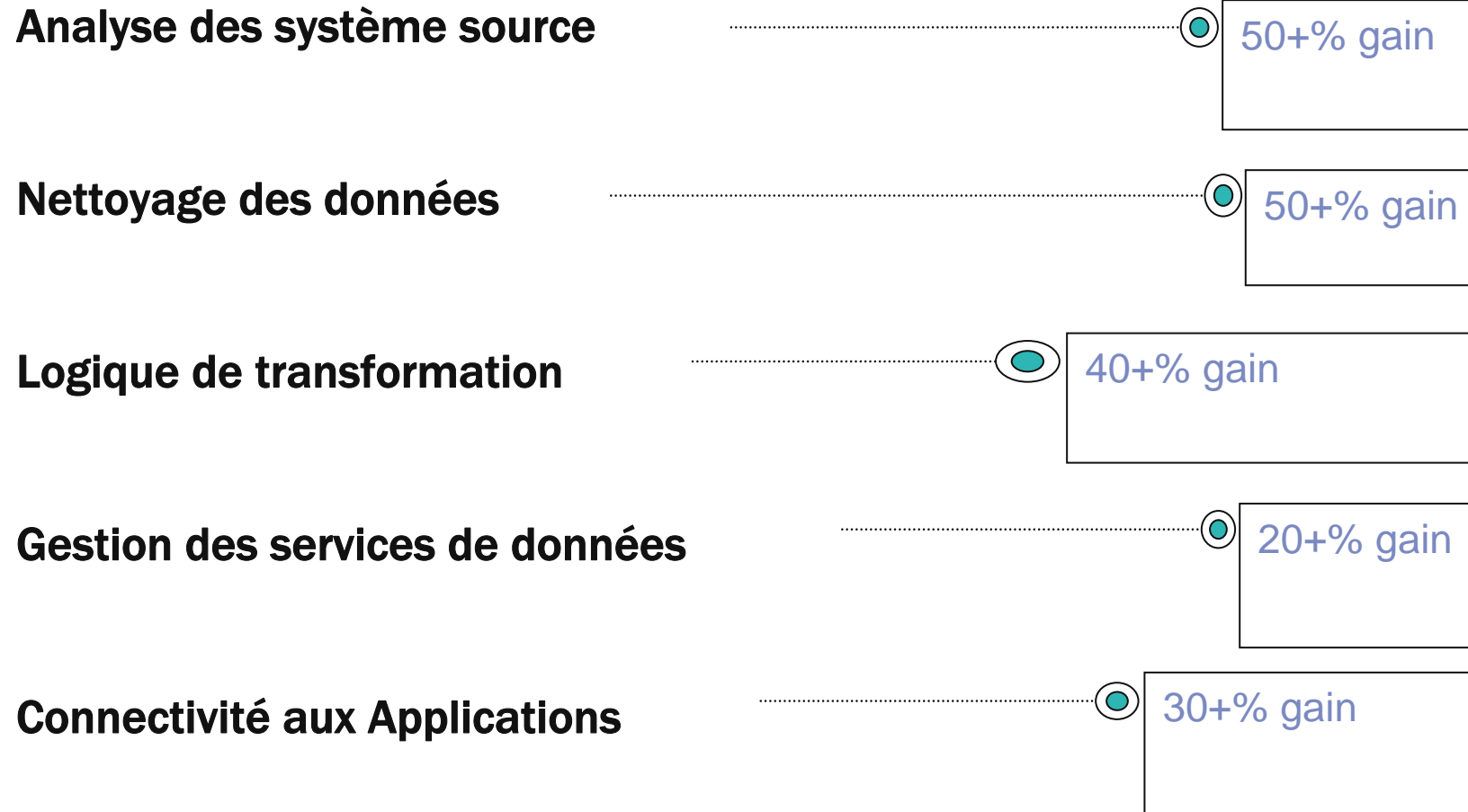
- Fusion du Manager dans le Designer
- Designer : gestion par arborescence + Référenciel
- Analyse d'impact graphique
- Différences entre jobs
- Parameters Sets
- Data Connection
- Nouveaux stages : Range Lookup, SCD, CFF
- SQL Builder pour de nouvelles BDD
- Procédures stockées avec support étendu
- + de Support des versions SGBD
- Composants natifs pour les composants de fédération (EII)
- SFTP Enterprise Stage (Secure data transmission)
- Connecteur natif iWay Enterprise
- Nouveau Connecteur SAP R/3
- Les Rich Connectors (ODBC, etc.)
- Gestion de la sécurité
- Analyse des performances
- Estimation de ressources
- Documentation technique
- Etc.

### Accéder aux modules additionnels

#### d'Information Server v8

- Analyse & Cartographie des données
- Gérer la Qualité des données au sein de Jobs DataStage
- Ouverture SOA en publiant des Jobs DataStage sous forme de Services.
- Capacité de Fédérer/Virtualiser (EII) l'accès a des sources hétérogènes de données.
- Gestion des méta données (techniques, runtime & métiers) globale et intégrées
- Traçabilité des méta données

# Les avantages sont réels



<sup>1</sup> Compared to hand coding – gathered from IBM project studies



# *Questions & conclusion ?*



**ON DEMAND BUSINESS™**