

PowerCenter et ses options

Edouard Guérin

Notre proposition de valeur

Améliorer les décisions

Supporter une fusion, une acquisition

Fidéliser les clients

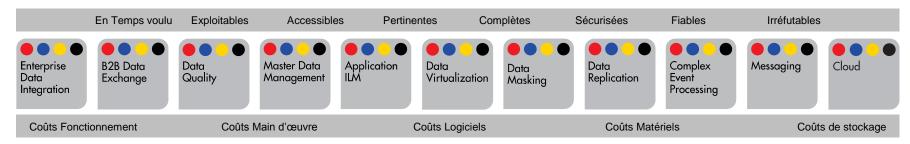
Faciliter la gestion des partenaires

Préserver la cohérence, la conformité

Assurer la confidentialité des données

Limiter les risques

AUGMENTER LA VALEUR DES DONNÉES



RÉDUIRE LES COÛTS DES DONNÉES















interactions

PowerCenter

Les éditions

Standard **SE**

- Moteur PowerCenter
- Profiling
- 2 connecteurs PowerExchange
- Data Services
 (non production)
- Data Analyst
 (5 utilisateurs)

Advanced



- Standard Edition
- Versioning
- Metadata Manager
- Business Glossary
- Data Analyst
 (10 utilisateurs)

Premium Pl



- Advanced Edition
- Monitoring Option
- Data Validation Option
- Data Analyst (20 utilisateurs)

PowerCenter

Les éditions

Standard

- Moteur **PowerCenter**
- **Profiling**
- 2 connecteurs PowerExchange
- **Data Services** (non production)
- **Data Analyst** (5 utilisateurs)

Advanced

- Standard Edition
- **Scaling Option** (HA. Load Balancing, Partitioning, Grid)
- Versioning
- Metadata Manager
- **Business** Glossary
- **Data Analyst** (10 utilisateurs)

Premium PE



- Advanced **Edition**
- **Monitoring** Option
- **Data Validation** Option
- **Data Analyst** (20 utilisateurs)

Realtime Package

- Advanced XML Opt.
- **Data Services** WebService Endpoint
- PowerCenter RT Opt.
- PWX for FAI
- **PWX for WebServices**
- PWX for CDC (2)

Monitoring Option

- **Proactive Monitoring** for PC Operations
- **Proactive Monitoring** for PC Governance

Scaling Option

- **Enterprise Grid** (HA, Partitioning, Load Balancing, Grid)
- **Pushdown Optimization**

Advanced Transformation

- **Unstructured Data Option** Excel, PDF, Word, **PowerPoint**
- **XMAP**

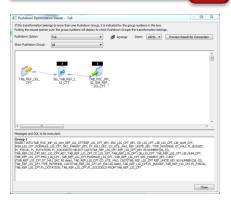
© 2015 Informatica Corporation – www.informatica.com

PowerCenter: optimiser les performances

Pushdown Optimization

Déportation des traitements en source, cible ou les deux. Dans ce cas, le moteur PowerCenter tire partie de la puissance de la base de données. La requête SQL est générée suivant les transformations demandées.





Partitioning

PowerCenter parallélise les traitements sur N partitions. Chaque partition peut être configurée suivant le volume, une clef ou une plage de valeurs. Le *« database partitioning »* utilise le parallélisme natif de la base de données.





Grid

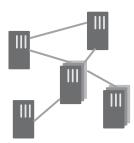
Le Grid distribue les traitements sur N nœuds. Cette architecture parallélise (physiquement) les traitements sur un ensemble de serveurs. Le partitioning peut, bien entendu, être activé en parallèle.



Change Data Capture

La lecture des « archive logs » permet de ne lire que les changements de la base de données. Cette solution réduit les temps de chargement tout en, éventuellement, traitant les données en temps réel.

Realtime Package

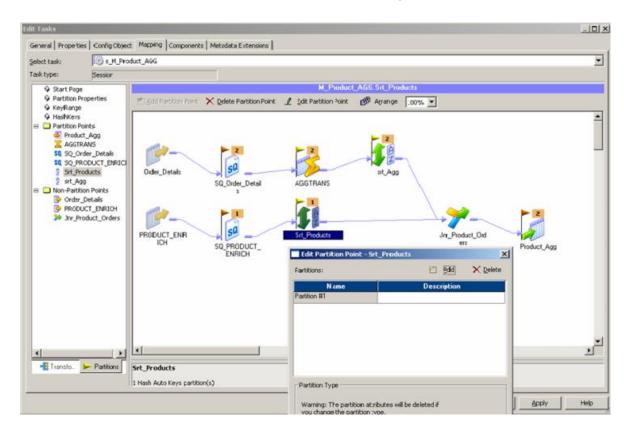




PowerCenter: quelques métriques

- PowerCenter traite, en moyenne, 10 Go de données par heure et par cœur de CPU
- Le partitioning et/ou grid permet diviser les temps de traitement (suivant l'infrastructure : nombre de CPU, réseau, I/O, etc.)
- PowerCenter utilise le maximum de l'architecture mise en place.
 A titre d'exemples :
 - 320 millions de lignes : 8,37 minutes (Oracle → Oracle distants),
 4 partitions, 4 cœurs de CPU / (160 000 lignes / secondes)
 - 4,9 millions de lignes : 11 secondes (Oracle → Oracle distants),
 Data base partitioning : 4 partitions (445 400 lignes / secondes)

PowerCenter: partitioning



- Hash user-key
- Round-robin
- Database partitioning
- Key range

PowerCenter: traiter les données complexes

Advanced XML

La lecture de fichiers XML complexes et / ou volumineux peut poser problème à PowerCenter. Advanced XML Option lève ces limitations pour traiter tout type de fichier XML : streaming, forts niveaux hiérarchiques, etc.

Realtime Package



Unstructured Data

L'intégration de données complexes ou non structurées n'est pas la vocation première de PowerCenter. Afin de traiter les fichiers Word, Excel, PDF, Logs, SWIFT, EDI, etc., Unstructured Data apporte une interface graphique pour convertir les données complexes.

Advanced Transformation



Social Media

L'extraction et l'exploitation des données sociales (réseaux sociaux) peuvent être chronophage si l'on ne maitrise pas les API Java ou WebServices. PowerCenter apporte ainsi des connecteurs prêt à l'emploi pour accélérer les projets : connaissance client, prospection, etc.

BSE

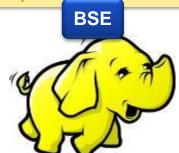






Big Data

Les architectures distribuées telles qu'Hadoop sont utilisées pour le stockage, l'archivage ou la distribution de calculs et traitements complexes, volumineux ou hétérogènes. PowerCenter BDE permet cela via ses connecteurs et son moteur adapté.



PowerCenter: accélérer les projets

Metadata Manager

Identifier et maintenir dépendances entre composants du svstème d'information est à la fois complexe et répétitif. collecte des métadonnées et l'analyse des impacts nécessite une solution de cartographie et de catalogue technique. AE

Business Glossary

Partager un dictionnaire commun permet de parler une même langue. MOA et MOE sont-ils en accord sur un même indicateur ? Quelle est la bonne définition ? La constitution d'un dictionnaire métier résout ce type de questionnement tout en favorisant l'échange.

File Edit Help | Legal Name | X | Legal Name |

Proactive Monitoring

supervision proactive permet de collecter en temps réel les évènements opérationnels (temps traitements, CPU, usage RAM, latence réseau, etc.) de sorte à anticiper les potentiels problèmes. Le monitoring est aussi apte à surveiller les bonnes pratiques.



Data Validation Option

La gestion des recettes et tests de non-régression prennent, généralement, 30% du temps projet. Informatica permet de réduire se temps via une interface dédiée aux tests : génération de cas et exécution historisée des résultats.





PowerExchange

Technologies standards	Batch	Delta / Temps réel
Serveurs mail (POP3 / IMAP)	•	
HTTP(S)	•	•
Java Message Service (JMS)	•	•
LDAP	•	•
ODBC	•	
Web Services	•	•
Bases de données	Batch	Delta / Temps Réel
DB2 UDB	•	•
Informix	•	
Microsoft SQL Server	•	•
MySQL	•	
Netezza	•	
Oracle	•	•
Sybase	•	
Teradata	•	

Applications	Batch	Delta / Temps réel
Hyperion Essbase	•	
Lotus Notes	•	
Oracle E-Business Suite	•	•
PeopleSoft	•	
SAP NetWeaver mySAP	•	•
SAP NetWeaver BW	•	•
SAS	•	
Siebel	•	
Messaging	Batch	Delta / Temps Réel
IBM MQ Series	•	•
Microsoft Message Queuing (MSMQ)	•	•
TIBCO	•	•
WebMethods	•	•
Cloud / SaaS	Batch	Delta / Temps Réel
salesforce.com	•	•

PowerExchange

Appliance	Batch	Delta / Temps réel
Teradata	•	
Netezza	•	
Greenplum	•	
Vertica	•	
Mainframe	Batch	Delta / Temps réel
ADABAS	•	•
Binary Files	•	
Datacom	•	•
DB2	•	•
IDMS	•	•
IMS	•	•
Tapes	•	
VSAM	•	•
Mid-Range	Batch	Delta / Temps Réel
ICL – IDMSX	•	•
DB2 – 400	•	•

Données structurées	Batch	Delt / Temps réel
	Daten	Deit / Tellips Teel
ACORD XML	•	•
MISMO	•	•
Metadata driven XML (deeply hierarchical and / or recursive)	•	•
Données semi-structurées	Batch	Delta / Temps Réel
EDI-FACT	•	•
EDI-X12	•	•
HIPAA	•	•
HL7	•	•
SWIFT	•	•
Données non-structurées	Batch	Delta / Temps Réel
HTML	•	
Fichier complexes	•	
MS Office (Word, Excel, PPT)	•	
Adobe PDF	•	
Emails, etc.	•	

Gouvernance des données

Gestion des métadonnées

Plateforme Informatica



Metadata Manager

Que signifie « profit » ? D'où vient la donnée ?

Développeur

Quelles données existent et quelles sont les règles métiers?



Data Analyst



Business User

Metadata

- Business
- Technique

Où la donnée est utilisée? Dans quel flux?

Qui possède la donnée?
Comment est-elle consommée?



Steward

Metadata Manager: collaboration MOA / MOE

Vision fonctionnelle

- Comprendre les données
- Communiquer avec un langage commun
- Partager l'information
- Relier les termes : synonymes
- Hiérarchiser les concepts

Dictionnaire de DONNEES

Vision technique

- Cartographier les systèmes
- Analyser les impacts
 - Identifier les sources et cibles
- Répertorier les dépendances
- Anticiper les changements

Catalogue de METADONNEES

GOUVERNER

Metadata Manager : collecter les métadonnées

Référentiel Metadata PowerCenter Bases de données Outils de modélisation Collecter les métadonnées Référentiel Metadata PowerCenter Outils BI Spécifique modélisation

Rechercher dans le catalogue

Analyser les impacts

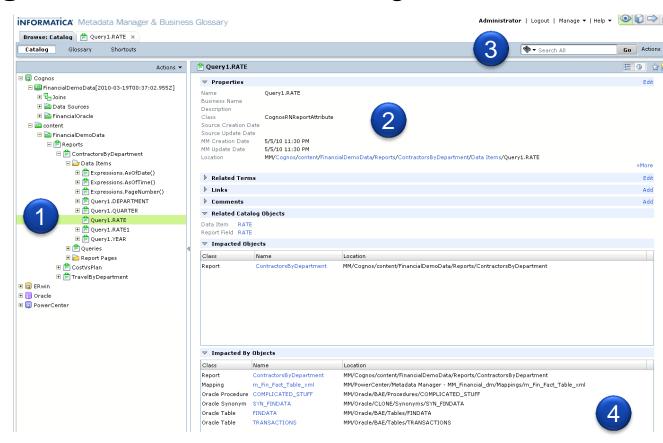
Assurer la traçabilité

Partager un langage commun

Enrichir et consulter un glossaire métier

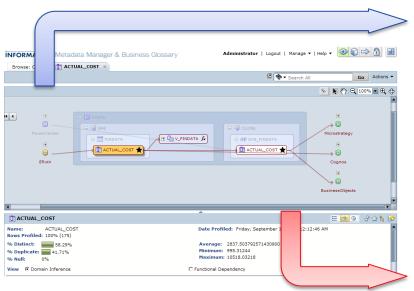
Metadata Manager: consultation du catalogue

- Navigation dans le catalogue. Lancer un data lineage se fait par simple clic droit
- Affichage des métadonnées : chaque objet a sa propre « fiche » consultable
- Possibilité de rechercher dans l'ensemble du catalogue (recherche avancée possible)
- 4) Liste des impacts



Metadata Manager: gestion des dépendances

Analyser les impacts

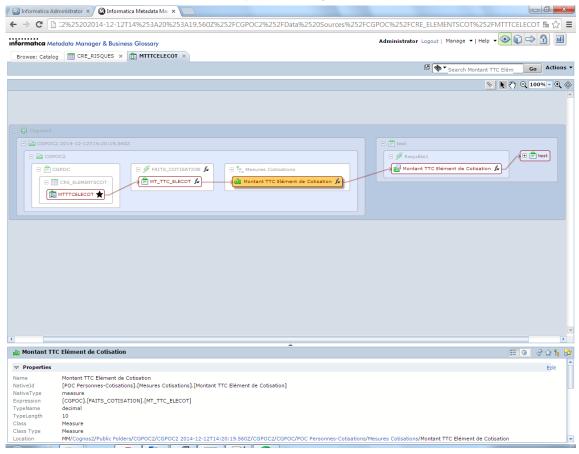


Identifier les dépendances en source et en cible

▼ Impact Summary - Downstream		
Class	Name	Location
Cognos Report	Actual Vs Planned Cost By Department	MM/Cognos/Public Folders/FinancialDemoData/Reports /Actual Vs Planned Cost By Department
Crystal Report	Cost vs Plan by Department	MM/BusinessObjects/Public Folders/Financial Reports/Crystal Reports/Cost vs Plan by Department/Cost vs Plan by Department
Document	CostvsPlanByDept	MM/BusinessObjects/Public Folders/Financial Reports/Desktop Intelligence/CostvsPlanByDept
Microstrategy Report	Cost vs Plan by Department	MM/Microstrategy/Financial/Public Objects/Reports/Cost vs Plan by Department
Web Intelligence Document	Cost vs Plan by Dept	MM/BusinessObjects/Public Folders/Financial Reports/Web Intelligence/Cost vs Plan by Dept

▼ Impact Summary - Upstream		
Class	Name	Location
Business Term	Actual Labor Cost	MM/BusinessGlossary/Costs/Actual Labor Cost
Business Term	Planned Cost	MM/BusinessGlossary/Costs/Planned Cost
Business Term	Planned Labor Cost	MM/BusinessGlossary/Costs/Planned Labor Cost
Cognos Report	Actual Vs Planned Cost By Department	MM/Cognos/Public Folders/FinancialDemoData/Reports /Actual Vs Planned Cost By Department
Crystal Report	Cost vs Plan by Department	MM/BusinessObjects/Public Folders/Financial Reports/Crystal Reports/Cost vs Plan by Department/Cost vs Plan by Department
Document	CostvsPlanByDept	MM/BusinessObjects/Public Folders/Financial Reports/Desktop Intelligence/CostvsPlanByDept
Mapping	m_Fin_Fact_Table_xml	MM/PowerCenter/Metadata Manager - MM_Financial_dm/Mappings/m_Fin_Fact_Table_xml
Oracle Synonym	SYN_FINDATA	MM/Oracle/CLONE/Synonyms/SYN_FINDATA
Oracle Table	FINDATA	MM/Oracle/BAE/Tables/FINDATA
Oracle Table	TRANSACTIONS	MM/Oracle/BAE/Tables/TRANSACTIONS
Web Intelligence Document	Cost vs Plan by Dept	MM/BusinessObjects/Public Folders/Financial Reports/Web Intelligence/Cost vs Plan by Dept

Metadata Manager: gestion des dépendances

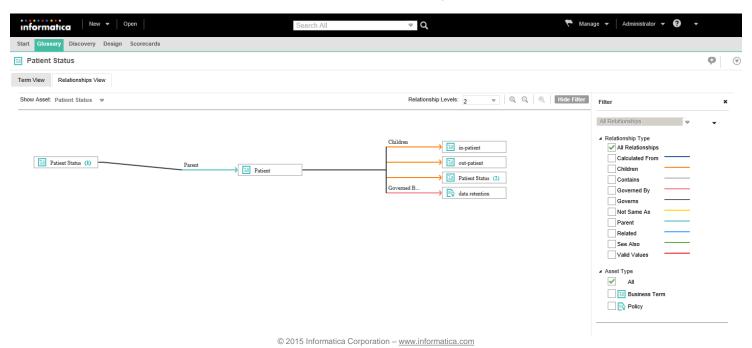


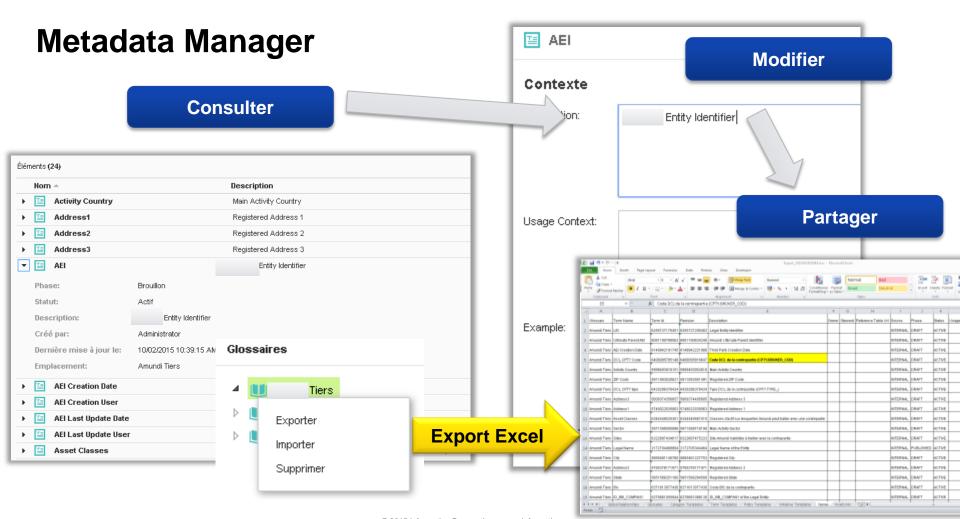
Les liens Cognos se font directement entre les mesures et les rapports avec leurs propres requêtes et indicateurs. Les outils BI tels que BO, Microstrategy et d'autres encore sont possibles.

Metadata Manager: consultation d'un dictionnaire métier

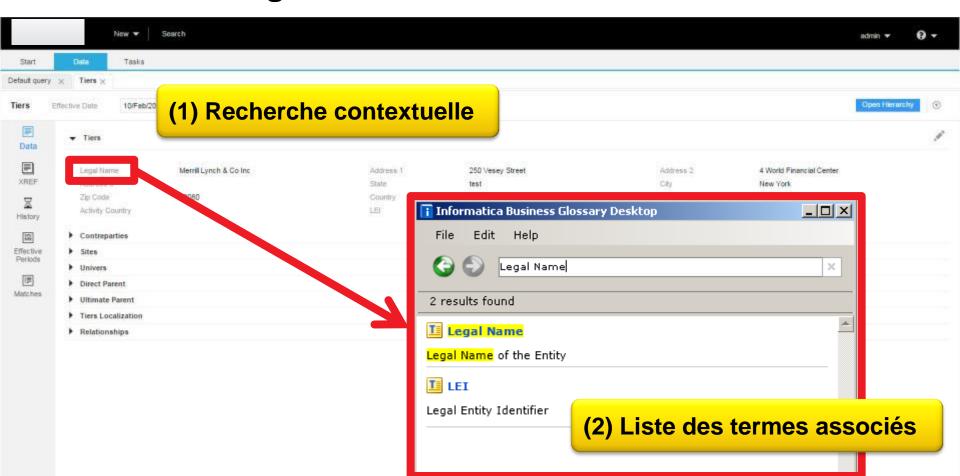
Le Business Glossary, dictionnaire métier, permet de :

- Partager les termes métiers / techniques
- Lier les termes métiers aux objets techniques (colonnes, reporting, etc.)
- Rechercher les définitions sur l'ensemble de l'écosystème





Metadata Manager: un utilitaire de consultation



L'intégration de données

Gestion des validations de données : DVO

Plateforme Informatica



La validation de données en général : le constat



Les validations de données se font généralement par scripts SQL et feuilles Excel.

Les résultats sont souvent partiels



La validation de données consomme environ 30% des efforts sur un projet d'intégration



Aucun temps n'est alloué aux validations de données des nouvelles mises en productions

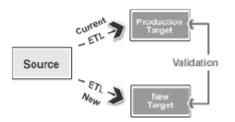
La validation de données par Informatica

- Solution basée sur & utilisant PowerCenter
- Génération automatique de mapping de tests
- Interface graphique simplifiée et dédiée aux utilisateurs en charge des tests
- S'appuie sur une liste de fonctions prédéfinies
- Multi-référentiels, multi-environnements

La validation de données par Informatica : cas d'usage

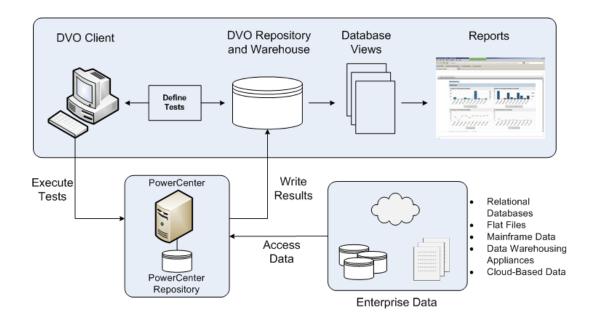
- Tous types de transformations
 - Sources → Cibles
- Tous types d'environnements
 - Recette ↔ Production
- Mises à jours
 - PowerCenter 8.X → 9.X
 - Base de données
 - Non régressions





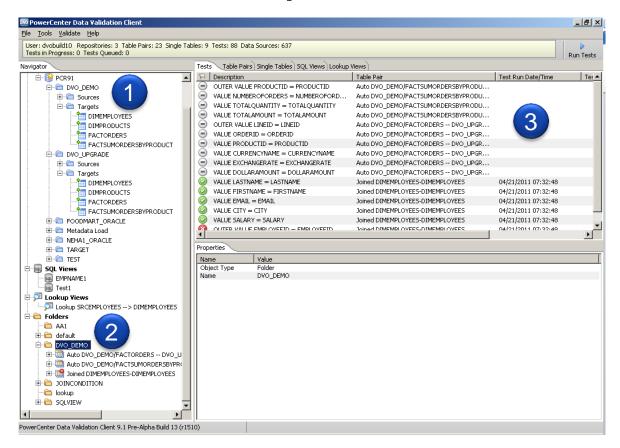


Data Validation Option : architecture



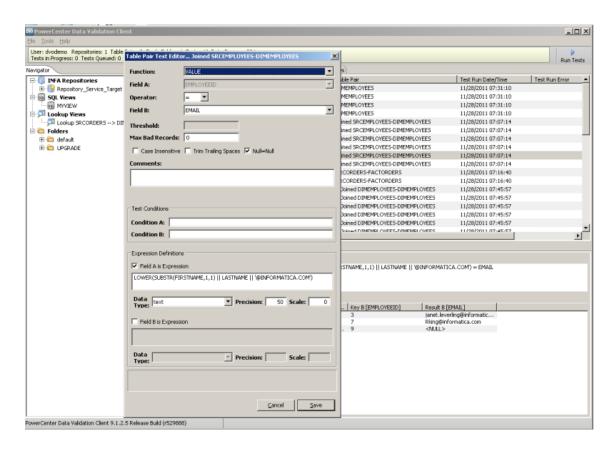
DVO est une interface supplémentaire à PowerCenter. Elle se greffe à l'architecture déjà en place. Un ensemble de tests génère un mapping PowerCenter de validation.

Data Validation Option: une interface simple



- L'utilisateur accède directement au référentiel. Les connexions sont préétablies.
- La création d'un jeu de test se fait graphiquement
- 3) Chaque test est visible dans la liste. Les tests sont exécutés uns à uns.

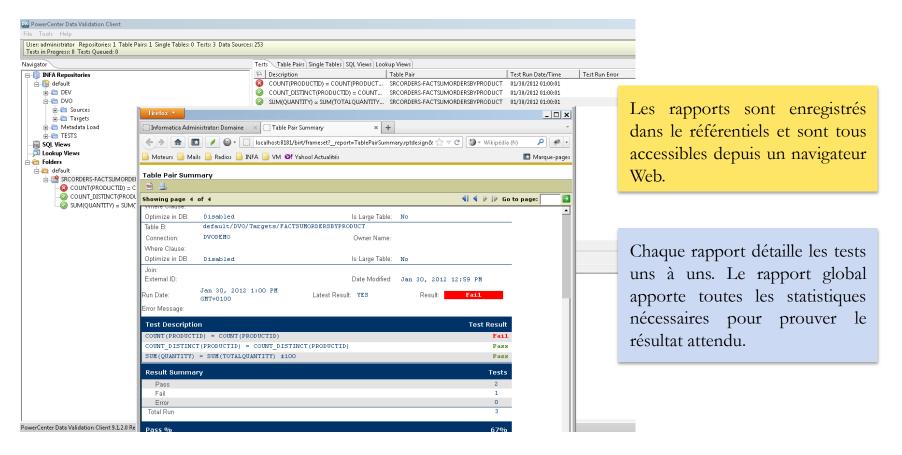
Data Validation Option: une interface simple



L'éditeur d'expression permet de créer tout type de test : sur valeur, sur approximation (avec seuil), est supérieur à, est inférieur à, etc.

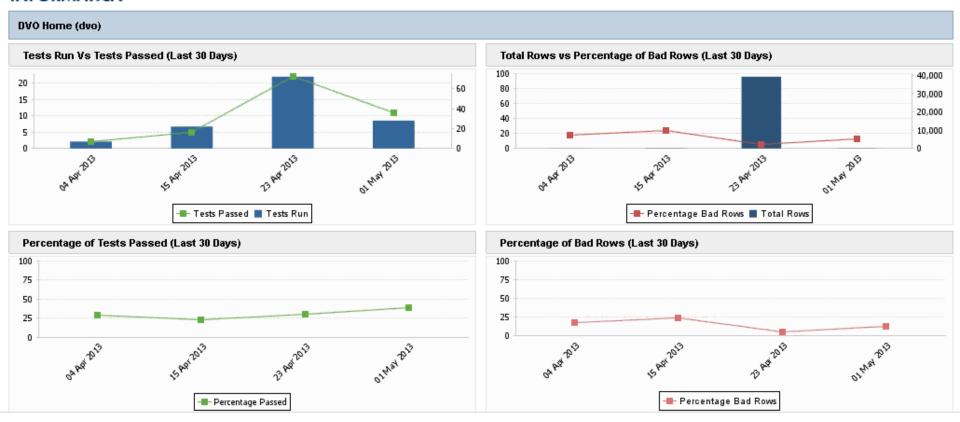
Les champs peuvent être transformés au besoin : gestion des dates, caractères, types, etc.

Data Validation Option : un rapport d'exécution

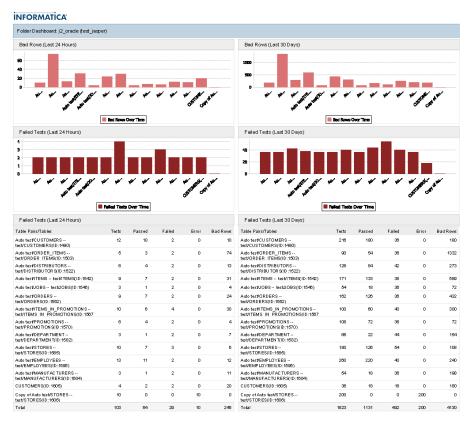


Data Validation Option: un dashboard

INFORMATICA'



Data Validation Option : une vision agrégée des tests



Questions?