



Livre blanc

Décisionnel
Solutions open source
Edition 2009

Version 3.0
Pour plus d'information : www.smile.fr
Tél : 01 41 40 11 00
Mailto : sd@smile.fr

PREAMBULE

Smile

Smile est une société d'ingénieurs experts dans la mise en œuvre de solutions open source et l'intégration de systèmes appuyés sur l'open source. Smile est membre de l'APRIL, l'association pour la promotion et la défense du logiciel libre.

Smile compte 290 collaborateurs en France, 320 dans le monde (septembre 2009), ce qui en fait la première société en France spécialisée dans l'open source.

Depuis 2000, environ, Smile mène une action active de veille technologique qui lui permet de découvrir les produits les plus prometteurs de l'open source, de les qualifier et de les évaluer, de manière à proposer à ses clients les produits les plus aboutis, les plus robustes et les plus pérennes.

Cette démarche a donné lieu à toute une gamme de livres blancs couvrant différents domaines d'application. La gestion de contenus (2004), les portails (2005), la business intelligence (2006), les frameworks PHP (2007), la virtualisation (2007), et la gestion électronique de documents (2008), ainsi que les PGI/ERPs (2008). Parmi les ouvrages publiés en 2009, citons également « Les VPN open source », et « Firewall est Contrôle de flux open source », dans le cadre de la collection « Système et Infrastructure ».

Chacun de ces ouvrages présente une sélection des meilleures solutions open source dans le domaine considéré, leurs qualités respectives, ainsi que des retours d'expérience opérationnels.

Au fur et à mesure que des solutions open source solides gagnent de nouveaux domaines, Smile sera présent pour proposer à ses clients d'en bénéficier sans risque. Smile apparaît dans le paysage informatique français comme le prestataire intégrateur de choix pour accompagner les plus grandes entreprises dans l'adoption des meilleures solutions open source.

Ces dernières années, Smile a également étendu la gamme des services proposés. Depuis 2005, un département consulting accompagne nos clients, tant dans les phases

d'avant-projet, en recherche de solutions, qu'en accompagnement de projet. Depuis 2000, Smile dispose d'un studio graphique, devenu en 2007 Agence Media Interactive, proposant outre la création graphique, une expertise e-marketing, éditoriale, et interfaces riches. Smile dispose aussi d'une agence spécialisée dans la Tierce Maintenance Applicative, le support et l'exploitation des applications. Enfin, Smile est implanté à Paris, Lyon, Nantes, Bordeaux et Montpellier. Et présent également en Espagne, en Suisse, en Ukraine et au Maroc.

Quelques références de Smile

Intranets et Extranets

Société Générale - Caisse d'Épargne - Bureau Veritas - Commissariat à l'Energie Atomique - Visual - CIRAD - Camif - Lynxial - RATP - Sonacotra - Faceo - CNRS - AmecSpie - INRA - CTIFL - Château de Versailles - Banque PSA Finance - Groupe Moniteur - Vega Finance - Ministère de l'Environnement - Arjowiggins - JCDecaux - Ministère du Tourisme - DIREN PACA - SAS - CIDJ - Institut National de l'Audiovisuel - Cogedim - Diagnostica Stago Ecureuil Gestion - Prolea - IRP-Auto - Conseil Régional Ile de France - Verspieren - Conseil Général de la Côte d'Or - Ipsos - Bouygues Telecom - Prisma Presse - Zodiac - SANEF - ETS Europe - Conseil Régional d'Ile de France - AON Assurances & Courtage - IONIS - Structis (Bouygues Construction) - Degrémont Suez - GS1-France - DxO - Conseil Régional du Centre - Beauté Prestige International - HEC - Veolia

Internet, Portails et e-Commerce

Cadremploi.fr - chocolat.nestle.fr - creditlyonnais.fr - explorimmo.com - meilleurtaux.com - cogedim.fr - capem.fr - Editions-cigale.com - hotels-exclusive.com - souriau.com - pci.fr - odit-france.fr - dsv-cea.fr - egide.asso.fr - Osmoz.com - spie.fr - nec.fr - vizzavi.fr - sogeposte.fr - ecofi.fr - idtgv.com - metro.fr - stein-heurtey-services.fr - bipm.org - buitoni.fr - aviation-register.com - cci.fr - eaufrance.fr - schneider-electric.com - calypso.tm.fr - inra.fr - cnil.fr - longchamp.com - aesn.fr - bloom.com - Dassault Systemes 3ds.com - croix-rouge.fr - worldwatercouncil.org - Projectif - credit-cooperatif.fr - editionsbussiere.com - glamour.com - nmmedical.fr - medistore.fr - fratel.org - tiru.fr - faurecia.com - cidil.fr - prolea.fr - bsv-tourisme.fr - yves.rocher.fr - jcdecaux.com - cg21.fr - veristar.com - Voyages-sncf.com - prismapub.com - eurostar.com - nationalgeographic.fr - eau-seine-normandie.fr - ETS Europe - LPG Systèmes - cnous.fr - meddispar.com - Amnesty International - pompiers.fr - Femme Actuelle - Stanhome-Kiotis - Gîtes de France Bouygues Immobilier - GPdis - DeDietrich - OSEO - AEP - Lagardère Active Média - Comexpo - Reed Midem - UCCIFE - Pagesjaunes Annonces - 1001 listes - UDF - Air Pays de Loire - Jaccede.com - ECE Zodiac - Polytech Savoie - Institut Français du Pétrole - Jeulin - Atoobi.com - Notaires de France - Conseil Régional d'Ile-de-France - AMUE

Applications métier

Renault - Le Figaro - Sucden - Capri - Libération - Société Générale - Ministère de l'Emploi - CNOUS - Neopost - Industries - ARC - Laboratoires Merck - Egide - ATEL-Hotels - Exclusive Hotels - CFRT - Ministère du Tourisme -

Décisionnel – Solutions open source

Groupe Moniteur - Verspieren - Caisse d'Épargne - AFNOR - Souriau - MTV - Capem -
Institut Mutualiste Montsouris - Dassault Systèmes - Gaz de France - CAPRI Immobilier -
Croix-Rouge Française - Groupama - Crédit Agricole - Groupe Accueil - Eurordis -
CDC Arkhineo

Applications décisionnelles

Bureau Veritas - Mindscape - Horus Finance - Lafarge - Brinks - Point P - Optimus Qualidata
- CecimObs - ETS Europe - Auchan Ukraine - CDiscount - Maison de la France Skyrock -
Institut National de l'Audiovisuel - Pierre Audouin Consultant - Armée de l'air - TVTrip -
Saint-Gobain Recherche - Xinek - Projectif - Companeo - MeilleurMobile.com

Ce livre blanc

Après avoir pris l'ascendant sur des marchés entiers, tels que la gestion de contenus, les portails, ou les frameworks de développement, l'open source est en train de gagner des parts de marché dans les outils d'aide à la décision, avec des solutions aujourd'hui très compétitives.

Comme les autres livres blancs publiés par Smile, cet ouvrage s'efforce de réunir :

- Une **approche générale de l'informatique décisionnelle**, ses concepts, ses champs d'application, ses besoins spécifiques.
- Un recensement des meilleurs outils open source dans le domaine du décisionnel.
- Une **présentation assez complète de ces outils**, de leurs forces, de leurs limites, de leur maturité et de leur aptitude à satisfaire des besoins opérationnels.
- Une présentation des offres de supports professionnels existantes en accompagnement de ces outils

Cette étude a été fondée sur plusieurs mois de travail de recherche, de déploiements effectifs, et optimisée en permanence par le biais des mises en œuvre opérationnelles pour nos clients.

Version 3

La première version de ce livre blanc remonte à Juin 2006. En 2007 nous avons publié une 2^{ème} version plus complète.

Cette version est riche des nouvelles versions des principaux éditeurs open source concernés à sa date de publication.

Ainsi nous avons une version 2.0 chez SpagoBI et Pentaho ainsi qu'une version 3.0 chez Jasper.

De plus nous ajoutons à notre panel la suite de la société Jedox comprenant un moteur M-OLAP.

Enfin, nous détaillons les offres de services des éditeurs open source.

SOMMAIRE

PREAMBULE..... 2

SMILE..... 2

QUELQUES REFERENCES DE SMILE **ERREUR ! SIGNET NON DEFINI.**

CE LIVRE BLANC 5

VERSION 3 5

SOMMAIRE..... 6

PRINCIPES 7

L'AIDE A LA DECISION 7

GENERATEUR D'ETATS *OU REPORTING* 7

ANALYSE A LA DEMANDE *OU ADHOC* 9

L'ANALYSE MULTIDIMENSIONNELLE *OU OLAP* 10

LA CONSOLIDATION DES DONNEES 14

LES PRINCIPES DE L'ETL 16

LE TABLEAU DE BORD 17

LE DATA MINING 18

TENDANCES..... 19

LES COMPOSANTS DECISIONNELS 20

PENTAHO DATA INTEGRATION : PDI (EX KETTLE)..... 20

TALEND 25

BIRT 29

JASPER REPORTS 32

PENTAHO REPORT DESIGNER..... 38

MONDRIAN 41

JPIVOT 41

FREEANALYSIS..... 43

PALO 47

JPALO 49

WEKA..... 50

LES SUITES DECISIONNELLES 52

PENTAHO 52

SPAGO BI..... 61

JASPER SOFT..... 65

JEDOX..... 70

SERVICES PROPOSES PAR LES EDITEURS..... 72

CONCLUSION..... 74

www.smile.fr

PRINCIPES

L'aide à la décision

Le chiffre d'affaires a baissé en octobre. Pour y remédier, des décisions sont à prendre. Mais pour prendre la bonne décision, il faut savoir pourquoi le chiffre d'affaires a baissé, et d'abord *comment* il a baissé. Dans quelle gamme de produits ? Dans quels pays, quelles régions ? Dans le portefeuille de clientèle de quels commerciaux ? Dans quel segment de distribution ? N'avait-on pas une baisse semblable en octobre chaque année ? Les questions sont nombreuses, auxquelles il faut savoir répondre avant de prendre la bonne décision.

On appelle « aide à la décision », ou bien « le décisionnel », ou encore « *business intelligence* », un ensemble de solutions informatiques permettant l'analyse des données de l'entreprise, afin d'en dégager les informations qualitatives nouvelles qui vont fonder décisions, qu'elles soient tactiques ou stratégiques.

Une entreprise moderne brasse dans son système d'information d'immenses volumes de données, mais bien souvent, du fait même de ces volumes trop importants, il est très difficile de *donner un sens* à ces données, de *comprendre ce qu'elles expriment* : des tendances sous-jacentes, des faiblesses ou des forces cachées, toutes choses que l'on doit connaître pour *prendre de bonnes décisions*.

Ainsi, les outils d'aide à la décision, avant d'aider à la prise de décision, aident d'abord à analyser les données afin *d'identifier les informations macroscopiques pertinentes cachées dans de grands volumes de données*.

Générateur d'états ou Reporting

Un *état*, en informatique, un rapport (*report* en anglais), est la présentation de données de manière synthétique et lisible, généralement à des fins d'impression.

L'état fait apparaître des tableaux de chiffres, en gérant entêtes, pieds de pages, sous-totaux, etc.

Un *générateur d'états* est un programme qui permet de définir un état selon ses besoins, avec un minimum de programmation, uniquement au moyen d'une interface interactive. Une fois défini, l'état peut être généré puis imprimé à échéance régulière.

Un état peut dépendre de *paramètres* : l'année, le mois, la région, la branche d'activité, etc.... Ainsi, un même état qui aura été défini une fois, pourra être généré avec différentes variantes, selon les valeurs prises par quelques paramètres.

Les paramètres de l'état pourront être renseignés par l'utilisateur selon ses souhaits, et il pourra également faire varier ces paramètres pour affiner sa recherche.

Dans d'autres cas, les paramètres sont définis automatiquement par des règles de gestion, par exemple pour adresser à chaque commercial l'état de ses propres ventes du mois. On parle ici de rapports en rafale ou de « bursting ».

On peut donc distinguer deux phases dans la génération d'un état :

- Une phase de conception initiale, qui requiert le plus souvent une expertise spécifique, et qui n'est pas entre les mains de l'utilisateur final ;
- Une phase de paramétrage et de production, qui ne requiert plus d'expertise et qui peut être mise entre les mains de l'utilisateur final.

Bien sûr, on aimerait limiter la première phase, qui est la plus coûteuse, et mettre directement des outils de conception entre les mains des utilisateurs finaux. Mais l'expérience montre que c'est rarement possible : même si l'apprentissage est possible, les utilisateurs ont en réalité d'autres priorités.

Toutefois en 2007, sont apparues au sein des solutions BI open source, des outils de reporting destinés aux utilisateurs finaux et dits « Adhoc ». Ceux-ci permettent aux utilisateurs finaux de réaliser rapidement et simplement des rapports simples, sans connaissance technique.

Analyse à la demande ou Adhoc

On peut distinguer deux modes de travail en décisionnel : le mode rapport statique et le mode interactif, avec une analyse qui répond interactivement aux demandes des utilisateurs.

Ils correspondent à des besoins différents, et parfois complémentaires :

- Obtenir une information récurrente, correspondant à une analyse prédéfinie, par exemple la décomposition des ventes par région ;
- *Chercher* une information en effectuant différentes analyses successives, les résultats de l'une amenant des questions nouvelles, qui demandent une nouvelle restitution de données.

Les outils peuvent être en partie les mêmes, mais il y a aussi des différences de fond :

- Les analyses récurrentes sont préparées en amont. Leur paramétrage peut nécessiter l'intervention d'experts. Une fois préparés, ils seront produits régulièrement, de manière quasi-automatisée.
- A l'inverse, les analyses interactives doivent être sous le contrôle de l'utilisateur final, qui n'est plus simplement consommateur d'analyse, mais en devient le principal acteur, sans être pour autant un expert des outils décisionnels. Cela implique que les outils intervenant dans ces analyses doivent être d'une grande simplicité d'emploi, autant dans les concepts fonctionnels que dans la configuration technique. Ils doivent notamment permettre de s'abstraire totalement de la technique, et en particulier du langage SQL de requêtage des bases de données, que les utilisateurs ne sont pas censés connaître.
- Enfin, l'analyse interactive ou adhoc requiert aussi d'excellents temps de réponse, pour permettre à l'utilisateur de tâtonner en direct en affinant son analyse.

L'analyse multidimensionnelle ou OLAP

L'analyse multidimensionnelle est l'un des modes d'analyse les plus courants dans le décisionnel.

Essayons d'en donner les grands principes.

On dispose de jeux d'informations élémentaires, généralement en grands volumes, chaque ligne d'information étant un évènement caractérisée par :

- Un identifiant unique
- Des attributs qualifiant l'information
- Des grandeurs portant une information quantitative

Pour être concrets, prenons l'exemple suivant, très classique, même s'il est très simplifié.

Considérons l'entité élémentaire qui est la *ligne de facture*. C'est souvent une entité clé dans une entreprise, l'information la plus fine dont on dispose par rapport aux processus de vente.

La ligne de facture appartient à une facture et porte sur la vente d'un produit, dans une quantité donnée, pour un prix donné à un client donné.

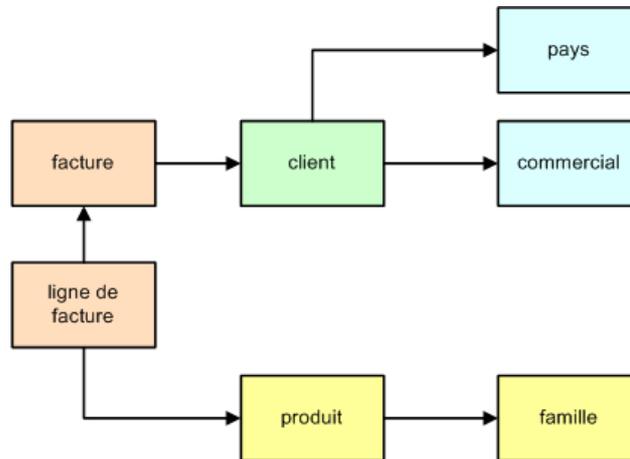
Elle porte en elle différents attributs : une indication de produit, une quantité, un prix unitaire, un prix total.

Au travers de la facture à laquelle la ligne appartient, on dispose d'autres informations : une référence de client, une date (année, mois, jour).

Sur le client lui-même, on possède d'autres informations : pays, région, type de client, secteur de métier, etc... Par ailleurs, le client est peut-être affecté à un collaborateur du commerce.

Nous nous arrêterons ici pour cet exemple, mais l'on voit bien que l'on peut pousser très loin cette *collecte d'information gravitant autour de l'entité la plus élémentaire qu'est la ligne de facture*.

Décisionnel – Solutions open source



Une première étape est donc la réunion de toutes les informations nécessaires à nos analyses. Ici, les informations sont par exemple :

- Date (année, mois, jour)
- Produit
- Famille de produits
- Client
- Pays du client
- Commercial
- Quantité
- CA

Les premières informations constituent les *axes* d'analyse potentiels, les deux dernières constituent les *grandeurs*.

Dans l'analyse multidimensionnelle, la modélisation relationnelle de départ n'est pas la plus pertinente, ni la plus efficace. On préfère généralement *dénormaiser* la base, c'est à dire travailler sur une table unique dans laquelle ont été rassemblées toutes les informations utiles.

Dans notre cas, on obtient :

Date	Produit	Segment	Famille	Client	Pays	Ccal	Quantité	CA
31/03/2007	12030991	Tondeuse	Jardinage	Castorama	France	Lepaul	50	50 000
31/03/2007	14030952	Perceuse	Outillage	Castorama	France	Lepaul	120	11 000
30/04/2007	12030993	Tondeuse	Jardinage	LeroyMerlin	Italie	Legrand	250	230 000
Etc....								

On remarque bien sûr qu'il y a énormément de redondance dans ce tableau, mais finalement il est plus utile ici de gérer de l'information redondante mais simple, et l'on ne s'occupe pas des problèmes de cohérence et d'intégrité qui sont du domaine des systèmes d'information.

L'étape suivante consiste à faire un premier niveau d'agrégation, c'est à dire à réunir certaines lignes.

Dans notre cas, on peut faire par exemple l'hypothèse que les données *ne seront pas utilisées au niveau de la référence produit*, mais uniquement *par segment*. Dans ce cas, on réunira toutes les lignes identiques pour la clé (date, segment, famille, client, pays, commercial), et l'on cumulera les grandeurs quantité et CA.

La dernière étape est celle de l'analyse multidimensionnelle proprement dite, qui consiste à sélectionner des axes d'analyse.

Parmi ces axes, on peut distinguer

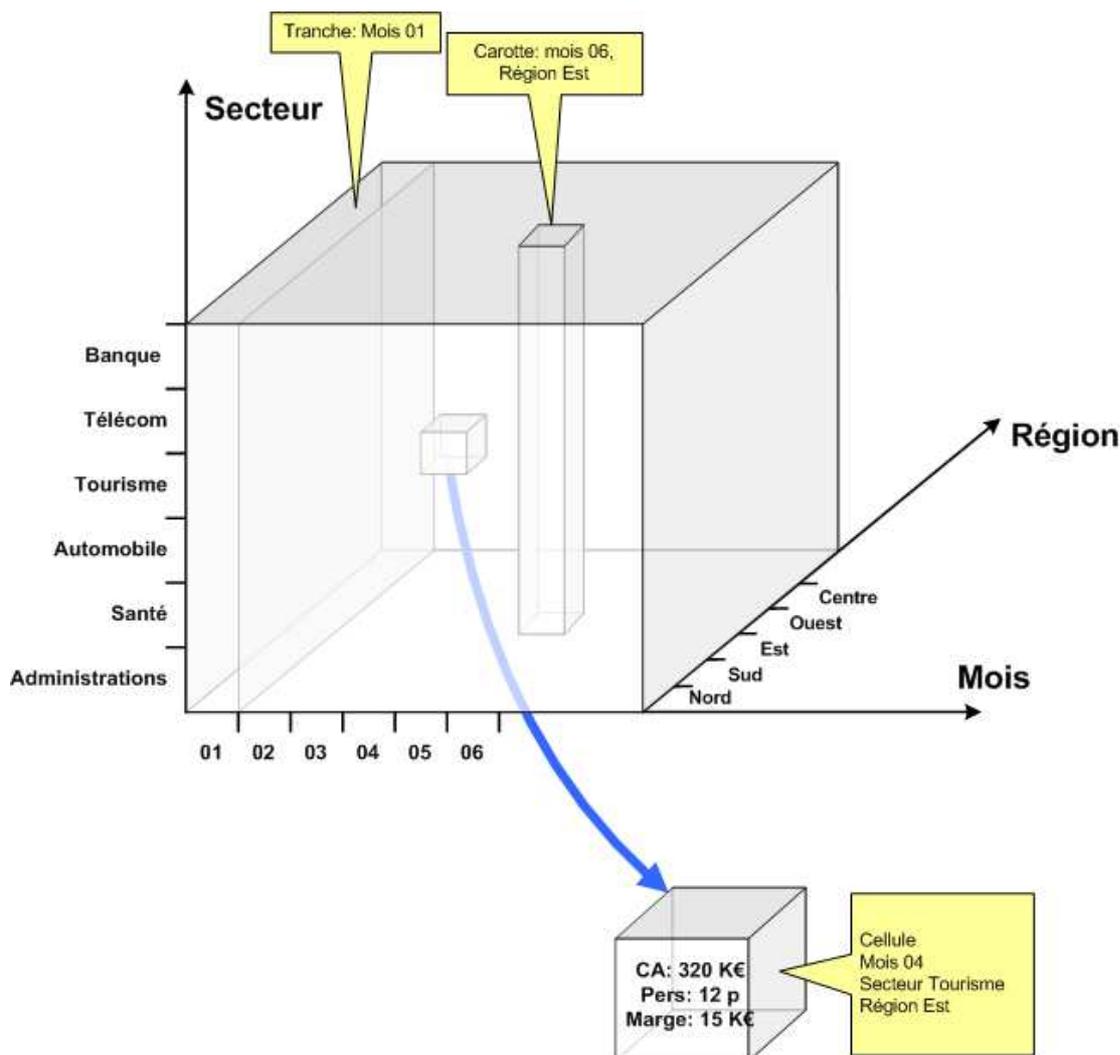
- Des axes à valeurs discrètes, ou discontinues, c'est à dire qui portent un nombre fini de valeurs, par exemple un code postal, un segment CSP.
- Des axes à valeurs continues, typiquement une date, un prix. On peut les ramener à un nombre discret de valeurs en définissant des *tranches* : tranches de prix, tranches d'âges.

On distingue également :

- Des grandeurs cumulables, par exemple un montant, un nombre d'items.
- Des grandeurs non cumulables, par exemple l'âge ou la date.

Les grandeurs cumulables sont celles qu'il est pertinent d'agréger, c'est à dire dont on peut calculer la somme, (ou la moyenne, ou d'autres fonctions mathématiques), pour un sous-ensemble de lignes, par exemple *pour chaque thématique*.

www.smile.fr



L'analyse multidimensionnelle consiste donc à :

- Définir les axes d'analyse que l'on utilisera, et l'ordre dans lesquels on les utilise. Par exemple : par région, puis par année, puis par vendeur, puis par gamme de produits. On n'utilisera pas toujours tous les axes possibles.
- Définir la ou les grandeurs qui sont étudiées, et éventuellement la fonction qu'on applique à ces grandeurs.

Chaque hiérarchisation des axes d'analyse correspond à une *question* que l'on se pose.

Par exemple une analyse *par année, par vendeur* vise à représenter l'*évolution* du chiffre d'affaires, puis à comparer les vendeurs chaque année écoulée. Une analyse *par vendeur, par année* vise avant tout à comparer les vendeurs (...).

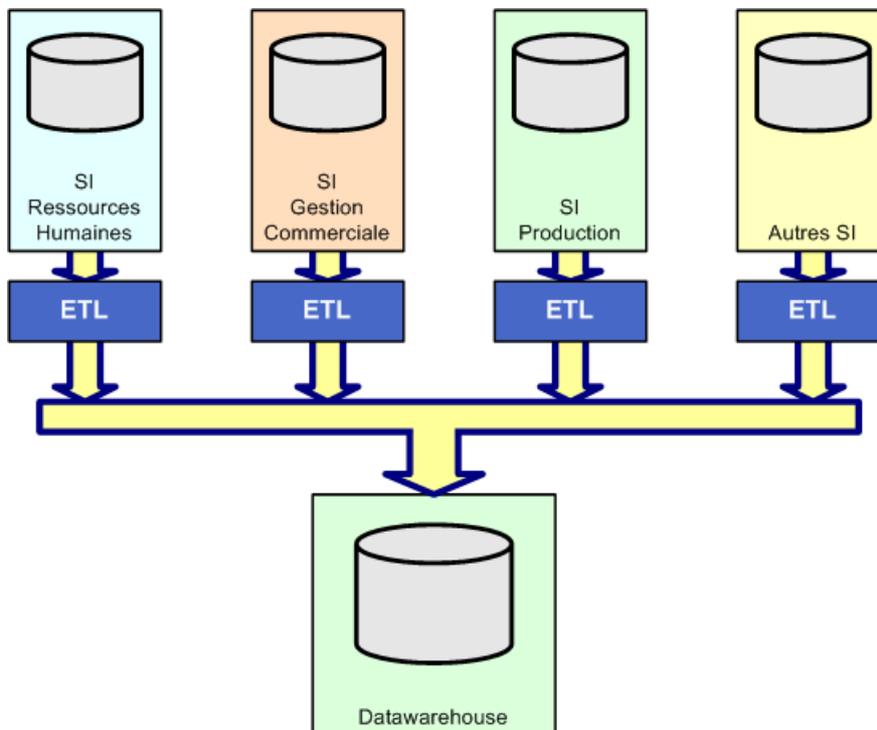
La consolidation des données

Principes

Les informations, les données, qui seront à la base du système décisionnel, proviennent souvent de multiples systèmes d'information de l'entreprise. Une application décisionnelle constitue le plus souvent un entrepôt de données, un « datawarehouse », en puisant les données de **plusieurs sources** dans l'entreprise.

Il y a de nombreuses raisons à cette étape de consolidation :

- La **centralisation** : il serait très difficile, et parfois impossible, d'accéder en temps réel aux données dans les différents systèmes où elles se trouvent : problèmes d'interconnexions réseaux, de débits, mais également de disparités de protocoles et d'interfaces,
- L'**unification** : le datawarehouse réunit les données dans un système unique, avec un référentiel commun, une modélisation unifiée, et des interfaces d'accès identiques. Il permet de créer des liens entre des données à l'origine hétérogènes.
- Les ressources informatiques : les applications décisionnelles peuvent être gourmandes en CPU, disque, mémoire, et les systèmes en place ne sont pas dimensionnés pour supporter ces nouveaux traitements.
- La spécialisation : les applications décisionnelles ont des besoins spécifiques qui ne pourront être satisfaits par les systèmes en place dans l'entreprise.
- Le **datawarehouse** est une base de données de consolidation, souvent une base relationnelle relativement standard, mais qui doit accueillir et manipuler de gros volumes d'information. Ce peut être une base de données open source (MySQL, PostgreSQL, ...) ou une base commerciale (Oracle, SQL-Server).



Le datawarehouse concentre de l'information décisionnelle issue de différents systèmes d'information de l'entreprise.

Organisation

Comme évoqué précédemment, cette nouvelle base de données au sein de l'entreprise n'est pas créée pour les bienfaits d'une application opérationnelle directement liée à l'activité de l'entreprise (ex : système comptable, base de données RH, base de données des commerciaux, ...). Le datawarehouse ne va stocker que les informations clefs de l'entreprise et ne sera dédié qu'aux requêtes d'analyse et de reporting.

L'entreprise pourra ainsi analyser ses données sans diminuer les performances de ces outils de production courants.

La base décisionnelle va être modélisée dans le but de faciliter les « requêtes ». On parle de modélisation décisionnelle en flocon ou en étoile.

Les tables ne vont être reliées que par un seul champ clef afin d'aider aux performances de requêtes et certaines

informations de type « liste de valeurs » vont être répétées de multiples fois dans les tables de dimension.

Comme toute modélisation de base de données, cette étape est cruciale dans la réussite d'un projet décisionnel. La modélisation doit être évolutive afin de facilement recevoir de nouveaux « datamarts » à l'avenir et assurer une réelle pérennité à l'entreprise en historisant un grand nombre d'informations.

Le datawarehouse se doit d'avoir une durée de vie plus longue que les outils de production. En effet, eux peuvent évoluer, voire être remplacés. Dans ce cas, le datawarehouse assure la mémoire de l'entreprise, et seules les interfaces de chargement de ce datawarehouse seront mises à jour.

Les principes de l'ETL

Pour alimenter le datawarehouse des différentes applications de l'entreprise, on utilise une gamme d'outils appelés ETL, pour « Extract, Transform, Load ». Comme le nom l'indique, ces outils permettent d'extraire des données à partir de différentes sources, de les transformer (format, dénomination), et de les charger dans la base de données cible, ici le datawarehouse.

Les transformations confiées à un ETL sont souvent simples, mais elles peuvent dans certains cas inclure des traitements procéduraux, de véritables programmes spécifiques.

Un ETL permet d'éviter la réalisation de programmes batch répétitifs, souvent semblables, dont il faudra également assurer la maintenance. Le principe est que l'intégration d'un nouveau flux de données ne requiert aucun développement, et s'opère par une simple configuration interactive : on choisit les éléments de données dans le référentiel source, on indique les transformations simples qu'ils doivent subir, et on précise la destination de la donnée dans le datawarehouse.

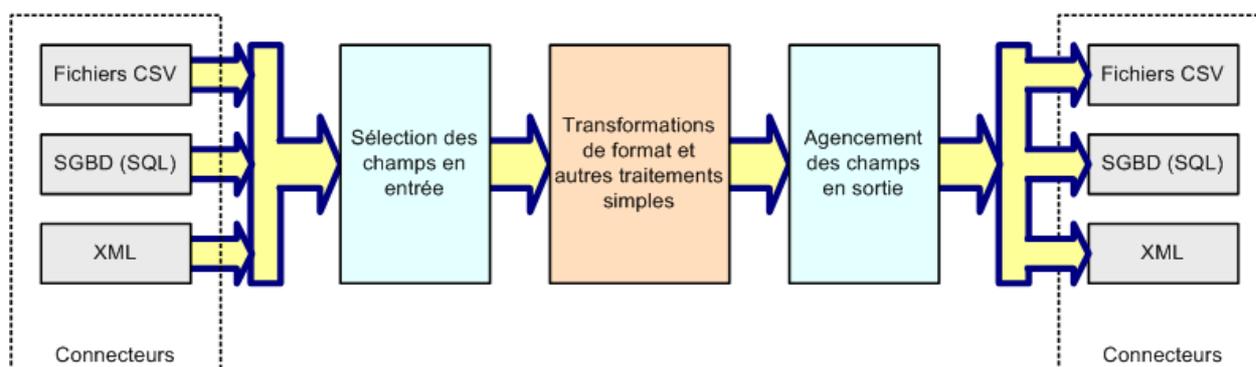
L'ETL peut prendre en charge différentes natures de sources de données, tant en entrée qu'en sortie, les principales étant bien sûr les SGBD relationnels, et les flux XML, mais il peut s'agir aussi de fichiers à formats fixes ou avec séparateurs (CSV).

Une fois qu'un flux d'extraction-transformation-chargement a été défini, il est généralement déclenché de manière régulière, ceci sous le contrôle d'un outil de planification de tâches, ou bien d'ordonnancement. (...)

Un ETL traite généralement des flux de point à point, c'est à dire entre une source unique et une destination unique.

L'ETL a vocation à travailler en différé, le plus souvent la nuit. Un datawarehouse enregistrant par nature une succession de « photos » de l'activité de l'entreprise, l'ETL va permettre d'alimenter le datawarehouse d'une « photo » supplémentaire. Une fois l'alimentation faite, les données ainsi chargées deviennent statiques et sont alors mises à la disposition des utilisateurs.

www.smile.fr



Le tableau de bord

Un tableau de bord n'est rien d'autre qu'une forme particulière d'état, de rapport.

On appelle tableau de bord un rapport qui est :

- Particulièrement *synthétique* : tout doit tenir sur une feuille A4, ou sur un écran d'ordinateur ;
- Le plus souvent *composite*, c'est à dire présentant plusieurs indicateurs, qui ensemble offrent une représentation complète de l'activité de l'entreprise ;

- Souvent également *personnalisé*, c'est à dire que chaque acteur consulte un tableau de bord *qui concerne son activité*.

On parle parfois de *Key Performance Indicators* (KPI), des indicateurs clés. Ce sont les indicateurs qui sont présentés dès la page d'accueil d'un portail décisionnel.

Selon les cas, le tableau de bord peut se suffire à lui-même, ou bien être le point d'entrée vers des analyses affinées, des tableaux de bord secondaires, ou bien des accès en *drill-down*.

On utilise parfois également le terme de *Executive Information System*, ou EIS, pour faire référence à un système décisionnel spécifiquement destiné à la Direction Générale, et donc présentant des informations stratégiques très consolidées.

Le data mining

Le *data mining* consiste à rechercher des informations statistiques utiles cachées dans un grand volume de données.

Dans le reporting de l'analyse multidimensionnelle, l'utilisateur sait ce qu'il cherche : il cherche par exemple la décomposition du chiffre d'affaires par région. Il ne sait pas encore quelle est cette décomposition, mais il sait du moins en quoi elle consiste, et qu'elle est une connaissance pertinente pour son travail.

Dans le data mining, l'utilisateur est à la recherche d'une information statistique cachée qu'il n'identifie pas encore : tendance, corrélation, similitude, etc. Typiquement, une analyse des tickets de caisse d'un hypermarché peut faire apparaître des corrélations entre achats de produits. Certaines auront une explication simple (rasoirs et mousse à raser), d'autres seraient insoupçonnées.

Il convient bien sûr de valider de tels résultats, car la corrélation observée peut ne pas traduire une relation de cause à effet, mais être fortuite, ou bien résulter d'une cause cachée, ayant l'un et l'autre phénomènes comme conséquences conjointes.

Les outils de data mining recherchent donc, de manière semi-automatisée, des corrélations, des règles statistiques, au travers de grands volumes de données, en s'aidant parfois

d'une restitution graphique de ces règles, qui fera apparaître par exemple des nuages de points sur une représentation axiale des données.

Tendances

Le décisionnel open source était jusqu'en 2006 constitué de composants écrits en Java dédiés à des fonctions spécifiques : édition de rapports, analyse OLAP, ETL... . Ces composants open source étaient écrits et maintenus par des communautés qui se sont progressivement professionnalisées.

La tendance est aujourd'hui clairement à la professionnalisation de l'open source décisionnel.

L'approche open source est devenu un « business model » très prometteur sur lequel les investisseurs misent, sous réserve que ce modèle économique soit piloté par un éditeur d'envergure mondiale.

Il est quand même remarquable de voir qu'en 2008, année économiquement très perturbée où les financements sont devenus très rares et très exigeants, à la fois Pentaho et Jaspersoft ont pu lever des fonds significatifs.

Ces éditeurs ont vocation à proposer des suites décisionnelles complètes sur lesquels ils ont la maîtrise de la technologie : le paysage décisionnel open source est ainsi entrain de se recomposer autour des grandes suites que nous analysons ci après.

A côté de ces grands éditeurs arrivent sur le marché des outils spécialisés plus jeunes. Nous en avons retenu deux, dont la maturité est évidente, dans cette analyse :

- ❖ Talend, dans le domaine de l'ETL
- ❖ Jedox, éditeur de Palo, dans un domaine non encore investigué par l'open source : une technologie M-OLAP au service de la saisie de données autant que des restitutions

LES COMPOSANTS DECISIONNELS

Avant de s'orienter vers la création de solutions décisionnelles complètes, les projets open sources se concentraient chacun sur un point bien précis du décisionnel.

Ainsi, les projets BIRT ou JasperReports permettent de générer des rapports, et les projets Mondrian et JPivot permettent de présenter des données sous forme multidimensionnelle. Ces projets étaient et sont encore destinés à être intégrés en tant que « composants » dans des développements spécifiques.

Les plateformes décisionnelles open source se basent sur ces composants déjà bien rodés et les intègrent de façon à constituer une **solution homogène**, dans laquelle toutes les fonctionnalités sont disponibles dans un cadre unique et rendues interopérables.

Dans cette partie, nous allons présenter les principaux composants décisionnels disponibles en open source, que l'on peut regrouper dans les catégories suivantes :

ETL : Pentaho Data Integration (Kettle), Talend Open Studio

Générateur d'états : BIRT, JasperReports et Pentaho Report Designer.

Analyse : Mondrian, JPivot, Palo et JPalo.

Data mining : Weka.

Pentaho Data Integration : PDI (Ex Kettle)

Présentation

PDI est un ETL open source qui permet de concevoir et exécuter des opérations de manipulation et de transformation de données.

Grâce à un **modèle graphique à base d'étapes**, il est possible de créer sans programmation des processus composés d'imports et d'exports de données, et de différentes opérations

de transformation, telles que des conversions, des jointures, l'application de filtres, ou même l'exécution de fonctions javascript.

PDI sous le nom de Kettle est devenu open source à partir de la version 2.2 et a rapidement intégré le projet de plateforme décisionnelle Pentaho (voir plus loin dans ce livre blanc) qui l'a renommé depuis en Pentaho Data Integration.

Aujourd'hui Pentaho Data Integration est disponible dans sa version 3.1. L'ergonomie a fortement été améliorée et de nouvelles étapes sont apparues.

Fonctionnalités

PDI permet de créer deux types de processus :

- Les *transformations* : traitements effectués au niveau d'une ou plusieurs bases de données comprenant des opérations de lecture, de manipulation et d'écriture.
- Les *tâches* : traitements de plus haut niveau, combinant des actions telles que l'exécution d'une transformation PDI, l'envoi d'un mail, le téléchargement d'un fichier ou le lancement d'une application. Il est possible d'exécuter des actions différentes en fonction de la réussite ou de l'échec de chaque étape.

PDI est également utilisable avec un référentiel partagé. Ainsi plusieurs développeurs peuvent utiliser des objets communs. Ce référentiel est stocké au sein d'une base de données relationnelle et le développeur peut régulièrement se connecter successivement à plusieurs référentiels.

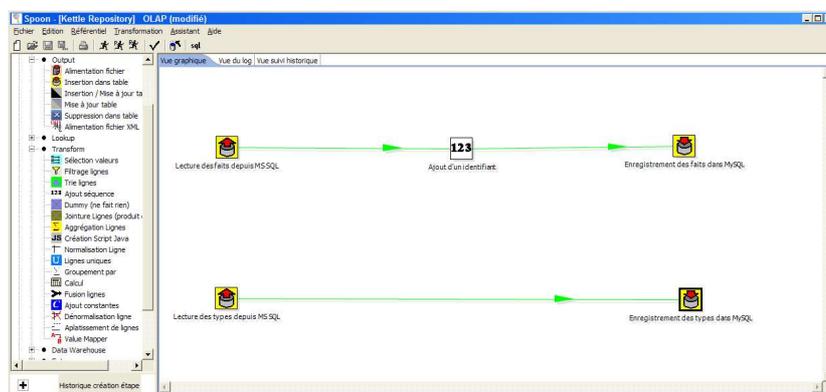
PDI peut se connecter sur un grand nombre de bases de données, dont Oracle, Sybase, MySQL, PostgreSQL, Informix, SQLServer et bien d'autres, et peut également utiliser des données provenant de fichiers textes, XML et Excel.

PDI dispose de :

- Détecteur de présence de fichier et lecture de répertoire
- connecteur en lecture ET écriture avec des cubes Molap PALO
- connecteur LDAP et lecture fichier LDIF
- connecteur Salesforce (lecture)

- liste des plug-ins disponibles sur : <http://wiki.pentaho.com/display/EAI/List+of+Available+Pentaho+Data+Integration+Plug-Ins>

Les transformations et les tâches sont créées par glisser-déposer des différentes étapes du processus. Des assistants sont disponibles à chaque étape et permettent par exemple de générer automatiquement les requêtes de lecture et de création de tables.



CREATION D'UNE TRANSFORMATION

On regrettera cependant l'absence de sauvegarde des différentes versions des jobs.

Les applications

PDI se compose de 3 applications :

- Spoon : environnement graphique de création et d'exécution de transformations et de tâches.
- Pan : application en ligne de commande permettant de lancer l'exécution d'une transformation donnée.
- Kitchen : application en ligne de commande permettant de lancer l'exécution d'une tâche donnée.

Pan et Kitchen sont utilisés pour planifier l'exécution des transformations et des tâches.

En installant seulement le très léger outil Pan sur plusieurs serveurs, les interfaces lancées par PDI vont pouvoir **être exécutées sur plusieurs serveurs simultanément**. On parle dans ce cas d'exécution en mode grappe. Vous avez ainsi le

moyen de maîtriser le temps de fonctionnements de vos interfaces lorsque vos « fenêtres » de nuit sont réduites.

Planification

Il est généralement souhaitable que des transformations ou tâches créées avec PDI s'exécutent périodiquement. C'est notamment le cas pour les tâches de synchronisation ou de création d'entrepôts de données.

PDI n'intègre pas directement de planificateur de tâches. L'exécution repose donc sur un système externe, tel que les planificateurs de tâches de Windows ou d'Unix.

En association avec la plateforme Pentaho et via la console de management de Pentaho, il est possible de planifier et suivre l'exécution de traitements PDI.

Utilisation

Dans le contexte de l'aide à la décision, PDI va servir à constituer un datawarehouse, en réalisant l'importation des données depuis différentes sources, leur consolidation et leur mise en forme.

L'utilisation de PDI présente plusieurs avantages par rapport à des solutions à base de scripts :

- La création de flux de données complexes est extrêmement simple
- La maintenance est beaucoup plus aisée car tous les flux de données sont visibles d'un simple coup d'œil dans une interface unique et centralisée.

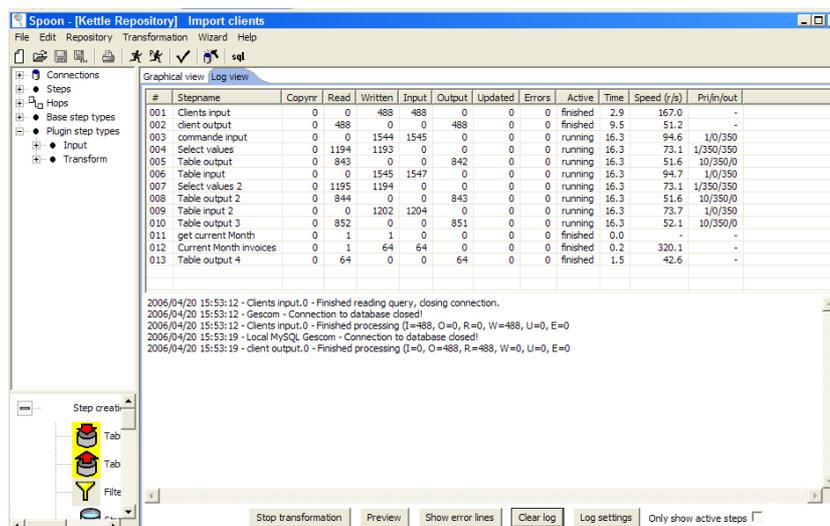
PDI pourra également trouver sa place dans tout projet non décisionnel qui a besoin d'exécuter périodiquement des traitements sur certaines données ou des synchronisations entre différentes bases.

Exemple

Détaillons les étapes nécessaires pour la création et le lancement d'une synchronisation entre 2 tables contenues dans des bases différentes :

1. Déposer un objet « Extraction depuis table » dans l'espace de travail.
2. Éditer l'objet « Extraction depuis table », choisir la source de données et la table concernées. Une requête SQL est générée automatiquement mais elle peut être modifiée librement.
3. Déposer un objet « Insertion dans table » dans l'espace de travail.
4. Lier ces 2 objets.
5. Éditer l'objet « Insertion dans table », choisir la source de données et le nom de la table. Kettle prend en charge la création d'une nouvelle table ou la modification d'une table existante, en utilisant les informations transmises par les étapes précédentes.
6. Lancer la synchronisation.

www.smile.fr



EXECUTION D'UNE TRANSFORMATION DANS SPOON

Conclusion

PDI est un outil efficace pour réaliser rapidement tout type de traitements sur des bases de données. Nécessaire dans la plupart des projets décisionnels, il sera également très utile dans tout projet utilisant des bases de données et demandant des opérations de synchronisation ou d'export.

PDI a l'avantage d'être performant, simple d'utilisation, entièrement open source et d'être interfacé avec la suite décisionnelle Pentaho présentée plus loin.

Talend

Présentation

Talend¹ est un ETL open source, développé par la société Talend, basée en France.

Talend est un ETL de type « générateur de code », c'est-à-dire qu'il permet de créer graphiquement des processus de manipulation et de transformation de données puis de générer l'exécutable correspondant sous forme de programme Java ou Perl. Ce programme Java ou Perl doit ensuite être déployé sur le serveur d'exécution et être exécuté.

Talend est l'ETL utilisé par les solutions SpagoBI et JasperSoft (où il est renommé en JasperETL).

Définition de processus

Les processus de manipulation de données sont créés en utilisant un modèle graphique à base de glisser/déposer. De nombreux types d'étapes sont disponibles pour se connecter aux principales bases de données ainsi que pour traiter différents types de fichiers (CSV, Excel, XML).

Talend dispose de fonctionnalités inédites pour un ETL open source :

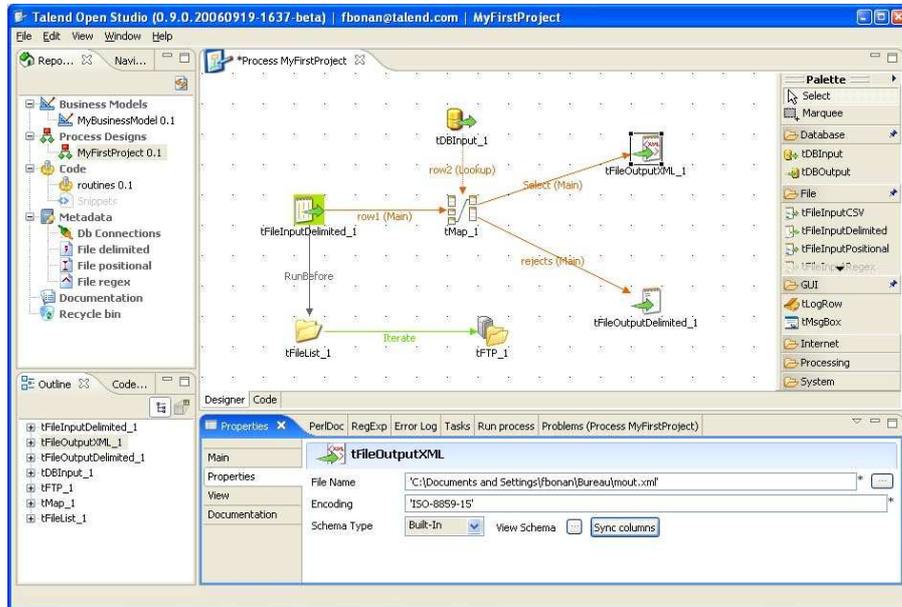
- les étapes ELT (Extract Load Transform), qui tirent parti de la base de données cible pour les différentes opérations, ce qui améliore grandement les performances au prix de possibilités plus limitées.
- Des connecteurs pour certaines applications de CRM (SugarCRM, SalesForces et CentricCRM), ce qui évite de manipuler les modèles relationnels de ces outils.

On notera que Talend facilite la construction des requêtes sur les bases de données en détectant les relations entre tables grâce aux clés étrangères et en proposant une interface « à la Access ».

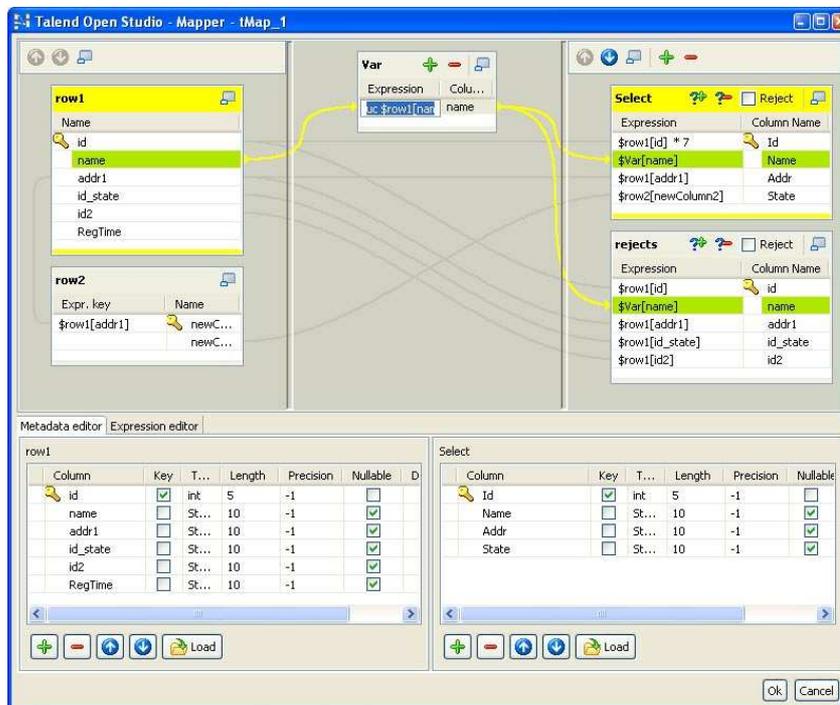
¹ <http://www.talend.com/>

Décisionnel – Solutions open source

Il est également possible d'ajouter simplement de nouvelles fonctions et composants afin de réaliser des processus plus complexes.



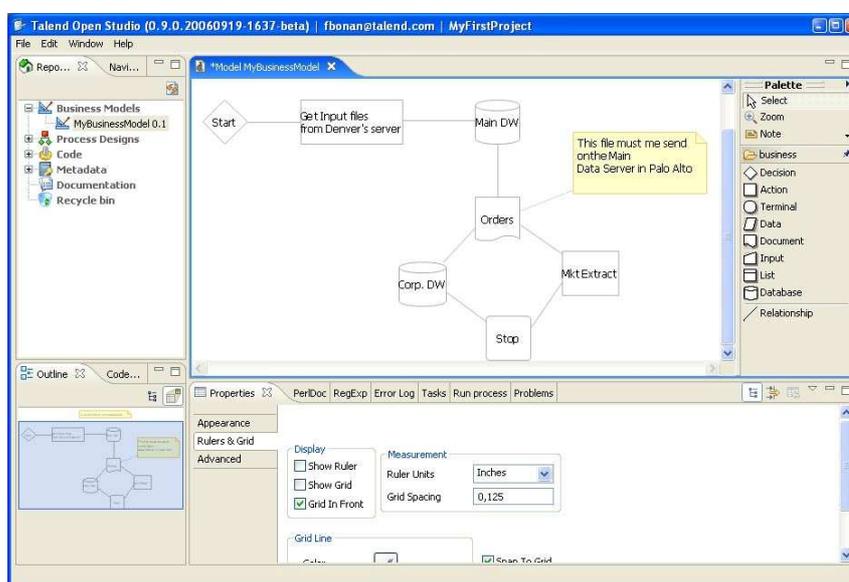
PROCESSUS



AFFECTATION ET MODIFICATION DE VALEURS

Modèles d'entreprise

Talend permet également de créer un modèle d'entreprise (Business Model) afin de modéliser les interactions entre les différents systèmes et bases de données. Les informations de connexion, les métadonnées ainsi que des documents peuvent être d'associés à chaque élément. Le modèle constitue alors une véritable documentation du système d'information.



MODELE D'ENTREPRISE

Ce mode est désormais complété avec un générateur de documentation technique permettant aux développeurs de gagner du temps dans l'étape souvent bâclée des spécifications techniques.

Gestion des contextes

Talend permet de **gérer les contextes** et de les appliquer aux transformations. Le même processus peut donc s'exécuter en environnement de développement, de test ou de production, avec à chaque fois l'utilisation des bonnes connexions aux bases de données et les bonnes métadonnées.

Déploiement des processus

Une fois le programme généré, celui-ci est installé par un administrateur sur la machine cible et son exécution est

Décisionnel – Solutions open source

planifiée en utilisant le service *cron* d'Unix ou les tâches planifiées de Windows selon le cas.

Avec la version TIS (Talend Integration Suite), est proposé une fonction « Distant Run » et « CPU Balancer » permettant à Talend de parfaitement s'intégrer dans des environnements professionnels industrialisés.

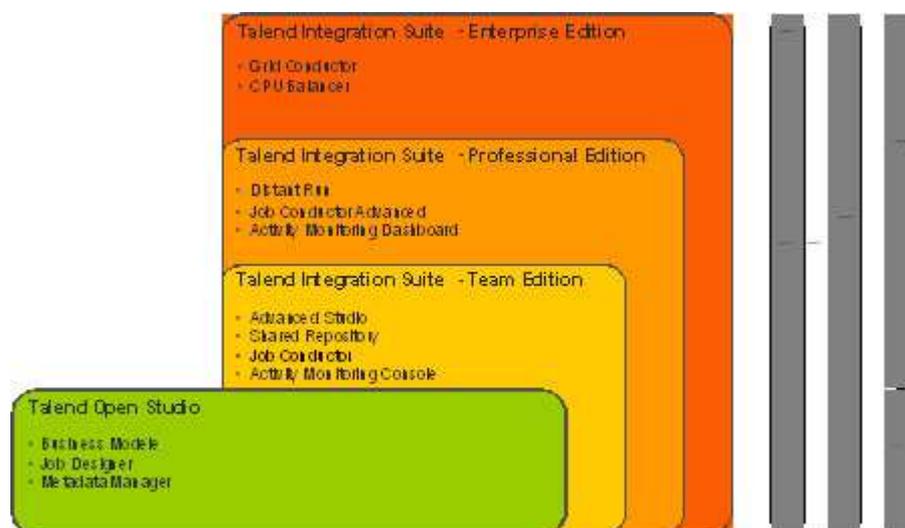
De par sa console de supervision en mode client Java ou en mode Web, les personnes chargées de l'exploitation des interfaces disposent de tableaux de bord synthétisant la bonne ou mauvaise exécution des traitements.

Conclusion

Talend est un produit complet. Talend a noué des partenariats avec de nombreuses sociétés éditrices de solutions décisionnelles ou de bases de données, ce qui renforce sa position sur le marché.

En 2008, Talend a complété son offre avec un nouvel outil « Talend Open Profiler » permettant de compléter l'activité ETL avec une gestion de la qualité des données entrantes. Cela permet par exemple de contrôler plus simplement l'unicité des données et de reconnaître automatiquement les données de même famille.

Notons que Talend propose une suite « Talend Integration Suite (TIS) », soumise à souscription annuelle, qui comprend des fonctionnalités très avancées comme la gestion des déploiements complexes, la supervision des exécutions et la gestion de référentiel partagés.



www.smile.fr

Comme PDI, Talend sera avantageusement utilisé dans des projets décisionnels mais trouvera encore plus sa place dans des projets d'urbanisation de systèmes d'information permettant d'uniformiser les modes d'échanges entre les différentes applications de l'entreprise.

BIRT

Présentation

BIRT² (The Business Intelligence and Reporting Tool) est un projet de la communauté Eclipse comprenant un générateur de graphiques, un générateur d'états et un environnement de conception.

Le projet a été initié par la société Actuate³, qui commercialise une offre de business intelligence intégrant BIRT et qui n'a adopté le modèle open source que pour BIRT.

Générateur d'états

Le moteur de BIRT est une bibliothèque qui permet de générer des états (paramétrés ou non) au format HTML, PDF, XLS, DOC ou PPT.

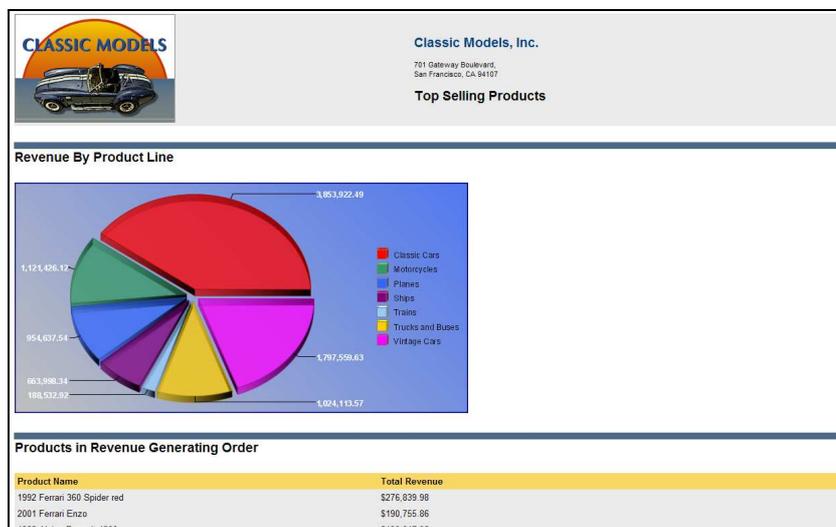
Ces états peuvent être complexes et contenir plusieurs tableaux, graphiques et images. Les données affichées peuvent provenir de bases et de requêtes différentes.

Le moteur de BIRT peut être intégré dans toute application développée avec le langage Java, que ce soit dans une application web ou dans une application de type « client lourd ».

² <http://www.eclipse.org/birt>

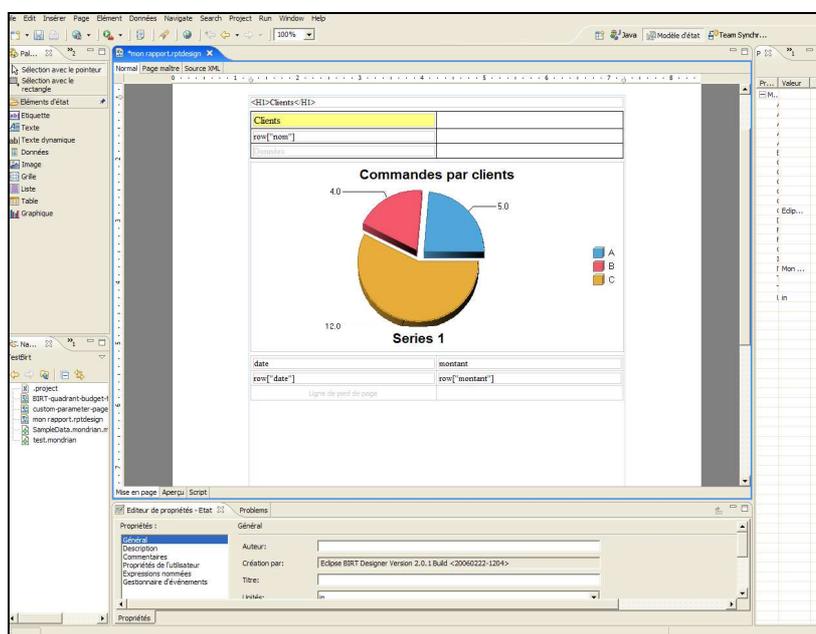
³ <http://www.actuate.com>

Décisionnel – Solutions open source



EXEMPLE DE RAPPORT BIRT

Conception des états



CREATION DE RAPPORT BIRT

BIRT dispose d'un outil de conception d'états, qui permet de créer et modifier les états de façon graphique, en déposant des éléments (tableaux, graphiques) dans un état vierge. Cet outil s'intègre sous forme de plug-in dans l'outil de développement

Décisionnel – Solutions open source

Eclipse⁴ mais peut être également utilisé comme une application autonome.

La définition d'un état étant simplement un fichier XML, il est également possible, mais beaucoup plus long et peu conseillé, de créer des états sans passer par cet outil.

La création d'un état passe par plusieurs étapes :

- La définition des sources de données : une ou plusieurs sources de données (bases de données compatibles JDBC ou fichiers texte) peuvent être ajoutées à l'état.
- La définition des jeux de données : un jeu de données est un ensemble d'informations destinées à être mises en forme par un tableau ou un graphique. C'est en fait le résultat d'une requête SQL plus ou moins complexe sur une source de données. L'utilisateur dispose d'un assistant pour faciliter la création de cette requête, et il peut de plus combiner différentes sources de données hétérogènes en les liant sur une clé commune.
- Le placement des éléments : les éléments de base (tableaux, listes, graphiques, images, ...) sont insérés dans l'état par « glisser-déposer ».
- La configuration : un jeu de données est associé à chaque élément et les paramètres de rendu sont modifiés.
- La prévisualisation : une option de prévisualisation permet de lancer la génération de l'état.

Audience

L'outil de conception d'états de BIRT s'adresse en premier lieu à des développeurs, car il est nécessaire de créer des jeux de données en utilisant des requêtes SQL.

Il est cependant tout à fait possible **de rendre BIRT accessible à des utilisateurs finaux** ayant un profil moins « technique », en utilisant certaines fonctionnalités qui permettent de contourner les étapes les plus complexes :

Dans un premier temps, il est possible de créer des **bibliothèques de ressources** contenant les éléments de base pour créer un nouvel état : les sources de données, la feuille de style de l'entreprise, et un ensemble de jeux de données.

⁴ <http://www.eclipse.org>

Ces éléments peuvent être incorporés dans un nouvel état par un simple glisser-déposer.

L'utilisateur pourra par la suite créer ses propres jeux de données en réalisant des jointures de jeux de données existants. Par exemple, si la bibliothèque contient les jeux de données '*clients*' (liste détaillée des clients) et '*ventes*' (liste des ventes incluant une référence vers le client, le produit, ...) , l'utilisateur pourra facilement créer un jeu de données '*ventes par client*' (liste détaillée des ventes pour chaque client). Il pourra ensuite ajouter des colonnes calculées et filtrer les résultats via un assistant.

Conclusion

BIRT est un outil de reporting très complet, qui permet de créer rapidement des états complexes. Il bénéficie du support de la très active communauté Eclipse.

L'interface de l'outil de conception des états est très intuitive et permet d'envisager de fournir cet outil aux utilisateurs afin qu'ils puissent concevoir de nouveaux états en fonction de leurs besoins.

Il y a encore quelques mois, on pouvait reprocher à BIRT de ne pas supporter certains types de graphiques et les tableaux croisés. Ce défaut est aujourd'hui corrigé. Non seulement les diagrammes de Gant, à bulles ou de différences sont disponibles, mais **BIRT permet maintenant de créer de façon intuitive et simple des cubes et des tableaux croisés**, à partir de jeux de données quelconques.

Jasper Reports

Présentation

JasperReports⁵ est un générateur d'états développé par la société JasperSoft⁶ et distribué sous une licence open source.

Plusieurs outils de conception sont disponibles pour créer de façon graphique de nouveaux états.

Les fichiers générés sont au format XML et peuvent également être créés et modifiés manuellement.

⁵ <http://jasperreports.sourceforge.net>

⁶ <http://www.jaspersoft.com>

Générateur d'états

Le moteur JasperReports permet la génération d'états au format PDF, HTML, XML, CSV, RTF, XLS et TXT. Il utilise JFreeChart pour générer les graphiques et peut être intégré dans toute application développée avec le langage Java.

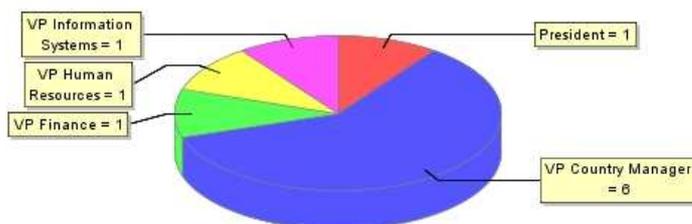
Il supporte, en plus des bases de données classiques, les serveurs d'analyse multidimensionnelle ce qui permet d'exploiter les possibilités du serveur Mondrian directement dans un état JasperReport.

www.smile.fr

Employment position

Role Senior Management

Position title	Pay type	Min scale	Max scale	Num
President	Monthly	25 000,00	85 000,00	1
VP Country Manager	Monthly	20 000,00	40 000,00	6
VP Finance	Monthly	17 000,00	35 000,00	1
VP Human Resources	Monthly	10 000,00	20 000,00	1
VP Information Systems	Monthly	12 500,00	25 000,00	1
TOT				10



JASPER REPORT

Conception des états

La conception des états se fait soit par description XML soit par outil graphique.

Les états sont décomposés en bandes dans lesquelles les éléments graphiques sont déposés. Chaque bande a un comportement spécifique et apparaît une ou plusieurs fois.

Un état exécute une itération sur un jeu de données principal. Certaines bandes sont affichées avant ou après l'ensemble des données de l'état, d'autres le sont une fois pour chaque élément du jeu de données.

		Titre de l'état
		Entête de la page
		Entête des colonnes
		Détails
		Fin des colonnes
		Pied de page
		Dernière page
		Page de résumé

Les différentes bandes disponibles sont :

- titre de l'état, affiché au début de la première page.
- entête de la page, affichée au début de chaque page.
- entête des colonnes, affichée avant les données.
- détails, répété pour chaque élément des données.
- fin des colonnes, affichée après l'ensemble données.
- pied de page, affichée en bas de chaque page.
- dernière page, affichée dans la dernière page.
- page de résumé, conclut l'état.

Pour créer des états plus riches, il est possible d'utiliser des jeux de données secondaires dans certains éléments, comme les graphiques et les tableaux ou d'insérer des états secondaires.

Les états JasperReports sont fortement liés au langage de programmation Java. Ils nécessitent une étape de compilation et des données fortement typées : il est par exemple

nécessaire de convertir en chaîne de caractères une donnée numérique avant de l'insérer dans un tableau.

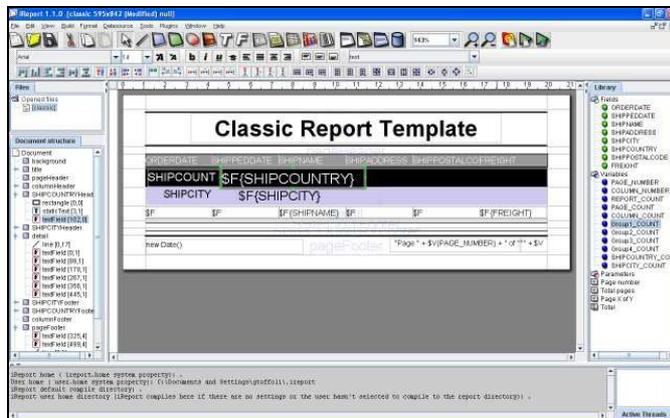
Il existe plusieurs outils de conception graphique pour les états JasperReports. Les plus importants sont iReport, Jasper Assistant et JasperPal.

iReport

iReport⁷ est l'outil de conception d'états officiel de JasperReports et se présente sous la forme d'une application Java dédiée.

Il supporte la totalité des fonctionnalités de JasperReports et dispose également d'une extension dédiée à l'administration de JasperIntelligence, la plate-forme décisionnelle de JasperSoft.

L'interface et le mode de fonctionnement des états JasperReports destinent principalement iReport aux développeurs. Comparée à BIRT, l'interface s'avère dans la version présente moins intuitive et moins efficace.



IREPORT

Jasper Assistant

Jasper Assistant⁸ est un outil de conception d'états qui s'intègre dans l'environnement de développement Eclipse. Il est distribué sous licence commerciale.

Le coût d'une licence est d'environ 130\$ et une version de démonstration est disponible sur le site de Jasper Assistant.

7 <http://ireport.sourceforge.net>

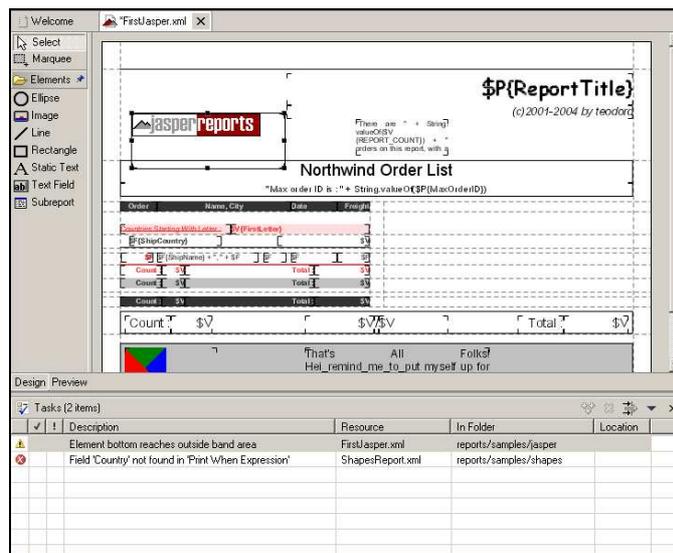
8 <http://www.jasperassistant.com>

Décisionnel – Solutions open source

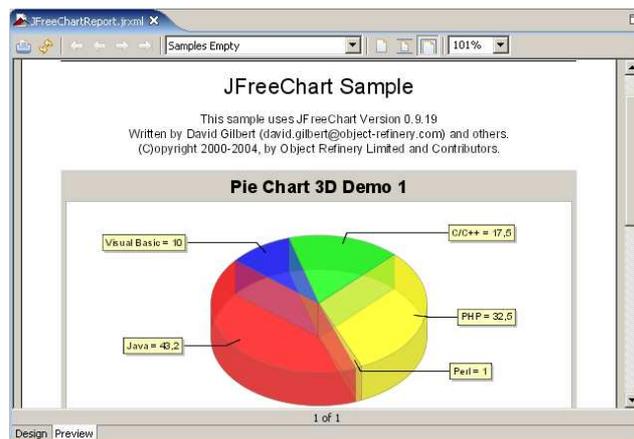
Le fonctionnement est semblable à celui d'iReport. En comparaison, l'interface bénéficie d'améliorations telles que l'édition des expressions et des formules qui est plus intuitive. En revanche, certaines parties telles que le paramétrage des graphiques sont plus complexes.

Au final, Jasper Assistant n'apporte pas de réelle plus-value par rapport à iReport, qui s'avère généralement plus complet.

www.smile.fr



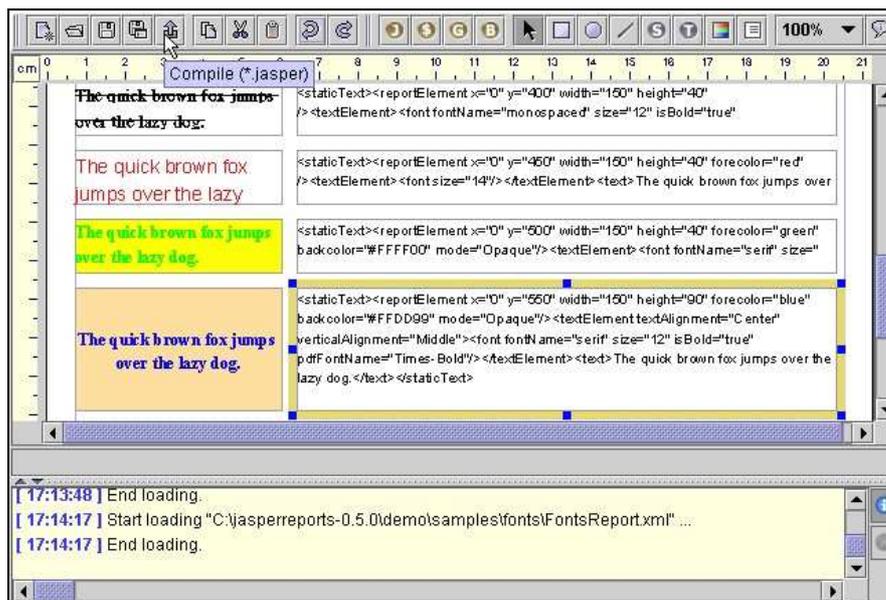
ÉDITION D'UN RAPPORT



PREVISUALISATION

JasperPal

JasperPal⁹ est un éditeur d'états destiné à être embarqué dans une application Java. Il est développé sous la forme d'un contrôle Swing et peut être utilisé dans toute application Java utilisant cet API.



COMPOSANT JASPERPAL

Plus limité que les autres éditeurs d'états, il peut cependant présenter un intérêt dans certains contextes spécifiques. Par exemple si l'on doit intégrer l'édition d'états dans un workflow non supporté par des outils existants. Dans ce cas une application qui englobera JasperPal et prendra en charge ce workflow devra être développée.

Conclusion

JasperReports est le moteur de génération d'états le plus répandu dans le monde open source. On constatera que toutes les plateformes décisionnelles permettent son utilisation. Son principal défaut est le manque d'intuitivité des différents éditeurs graphiques, ce qui limite leur utilisation aux développeurs.

⁹ <http://jasperpal.sourceforge.net>

Pentaho Report Designer

Présentation

JFreeReport a rejoint le projet Pentaho début 2006. Au fur et à mesure, le nom JFreeReport a été abandonné au profit de Pentaho Report Designer (PRD).

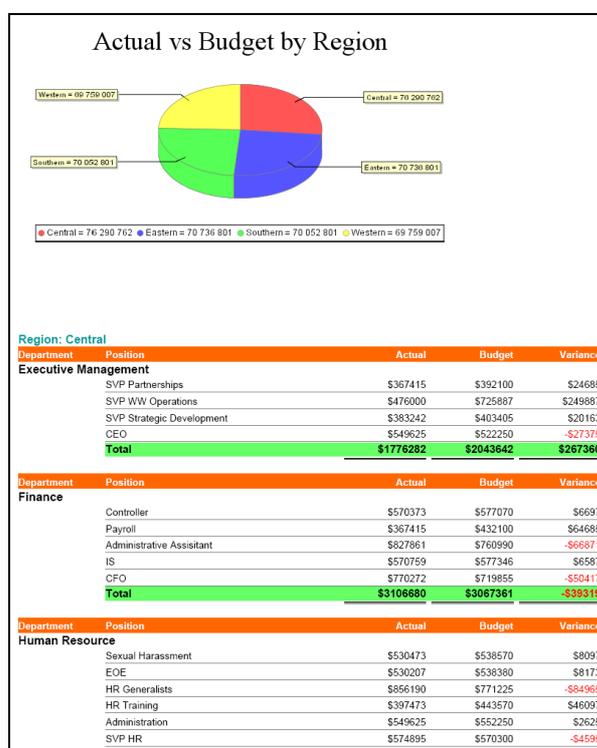
PRD permet de développer des rapports complexes et en association avec la plateforme Pentaho de les publier directement sur le serveur décisionnel.

Générateur d'états

PRD supporte les formats PDF, HTML, CSV, Excel et Texte. Très modulaire, il permet aux utilisateurs de développer en Java de nouveaux types de contenu suivant leurs besoins.

Ce moteur se base également sur des définitions d'états en XML et peut être intégré dans toute application Java.

www.smile.fr



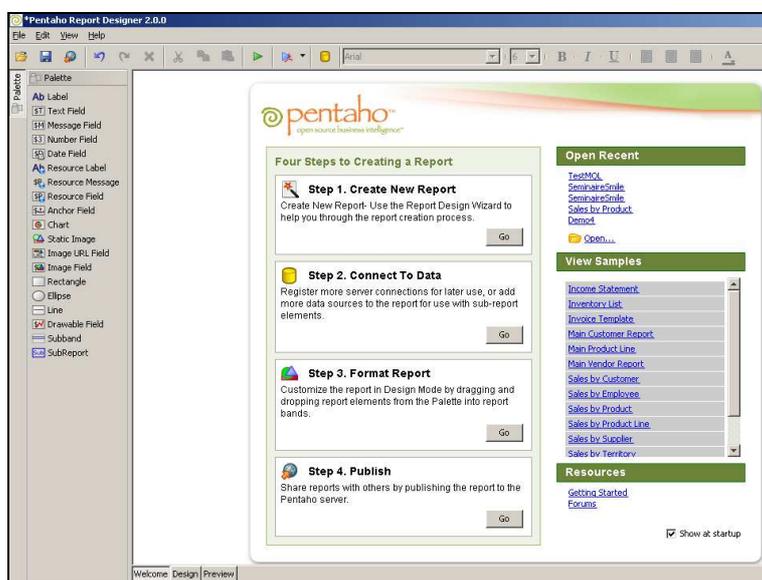
PENTAHO REPORT DESIGNER

Conception des états

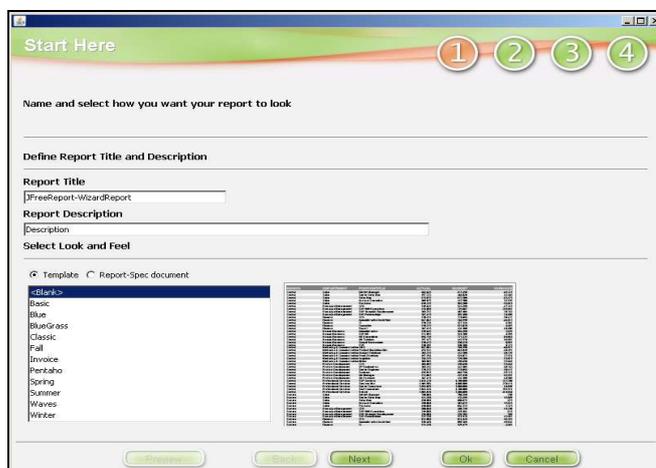
Les états PRD se structurent en bandes, de la même façon que dans JasperReports.

Pentaho Report Designer est un outil de conception graphique similaire à iReport. Avec la version 1.7 et aujourd'hui la version 2.0, PRD s'est considérablement enrichi.

Ainsi dès l'ouverture une page d'accueil guide l'utilisateur vers les étapes indispensables de création du rapport :



En 1^{ère} phase il est proposé un Wizard permettant d'établir une nouvelle connexion mais surtout désormais de pouvoir se connecter à un metadata (dictionnaires de données) de Pentaho.

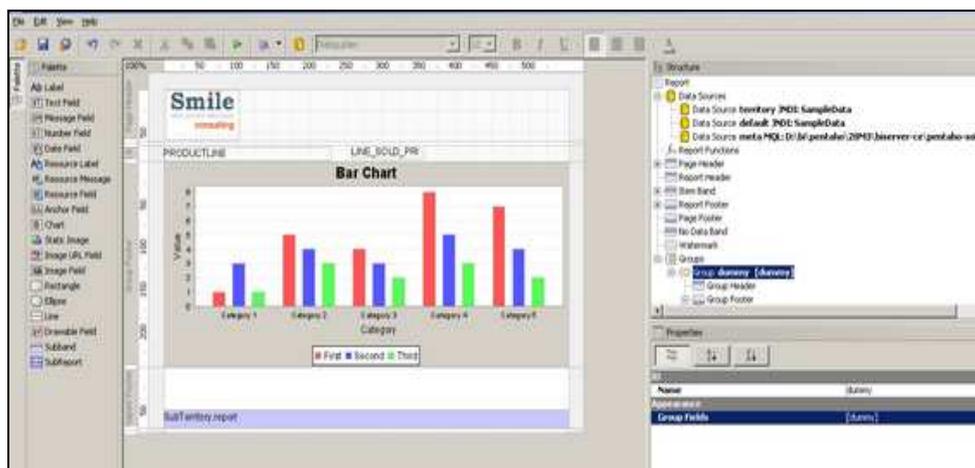


Il s'agit ici d'une avancée importante dans les outils de reporting open source et PRD a été le premier fin 2007 à permettre à un utilisateur de créer un rapport sans aucune connaissance SQL.

Dans la partie « Design » on dispose du même éditeur de formule que le tableur de OpenOffice, permettant de créer des champs calculés.

Plusieurs fonctions sont également disponibles afin de réaliser des opérations courantes comme le comptage d'occurrences dans un regroupement ou des fonctions de conversions de format. Ces fonctions sont présentées dans une bibliothèque par thème.

www.smile.fr



Conclusion

Pentaho Report Designer est un outil simple à manipuler, bien intégré à la suite décisionnelle Pentaho pour la gestion des paramètres ou la publication sur la plateforme web.

On regrettera cependant dans PRD l'impossibilité de créer des tableaux croisés sauf à pointer sur un schéma Mondrian existant. Toutefois cette fonction est annoncée dans la roadmap 2009.

Mondrian

Mondrian¹⁰ est un serveur OLAP (On Line Analytical Processing) ou « moteur OLAP » disponible sous licence open source.

Il fait partie de la catégorie des serveurs R-OLAP, c'est-à-dire qu'il accède à des données contenues dans une base relationnelle.

Mondrian exécute des requêtes utilisant le langage MDX, également utilisé dans Microsoft SQL Server. Ce langage permet de créer des requêtes dont l'équivalent en langue SQL nécessiterait un grand nombre de requêtes et des temps d'exécution beaucoup plus longs.

Ce serveur est utilisé avec les clients JPivot ou JPalo (présentés ci-après), outils qui proposent une interface graphique de consultation et manipulation des données.

Le projet Mondrian a maintenant rejoint le projet Pentaho sous le nom de Pentaho Analysis.

JPivot

Présentation

JPivot¹¹ est un client OLAP disposant d'une interface Web. Il permet de représenter un cube OLAP sous forme de tableau croisé multidimensionnel et d'effectuer les opérations classiques d'analyse (drill down, drill up, rotations, filtres ...) de façon interactive.

JPivot permet également d'afficher un graphique correspondant aux données présentées en tableau, qui est mis à jour au fur et à mesure de l'exploration.

Il est possible d'exporter tableau et graphique sous forme de fichier PDF imprimable ou de document Excel afin de réutiliser les données obtenues.

L'interface utilisateur peut être facilement modifiée via l'utilisation de feuilles de styles (CSS et XSL).

¹⁰ <http://mondrian.sourceforge.net/>

¹¹ <http://jpivot.sourceforge.net/>

Décisionnel – Solutions open source

www.smile.fr

		Measures			
Products	Region	Measures[0]	Measures[1]	Measures[2]	
-All Products[0]	-All Region[0]	1,015.22	780.70	890.02	
	All Region[0]				
	+Region[0]	873.77	971.56	1,102.23	
	-Region[1]	1,074.22	945.91	823.55	
	Region[1]				
	+City[0]	876.48	923.64	1,026.13	
	+City[1]	1,136.24	825.63	1,067.35	
	+City[2]	955.63	1,071.25	1,159.65	
	+City[3]	1,163.37	1,015.91	949.34	
	+City[4]	909.86	1,136.97	984.30	
	+City[5]	1,116.37	1,063.09	985.76	
	+City[6]	1,089.23	1,063.38	1,006.19	
	+City[7]	1,172.89	992.38	1,050.60	
	+Region[2]	983.33	834.35	1,131.19	
+Region[3]	1,231.73	1,041.51	1,026.49		
+Region[4]	1,043.67	1,167.71	903.53		
All Products[0]	+Category[0]				
	-All Region[0]	890.73	1,170.83	955.32	
	All Region[0]				
	+Region[0]	1,121.95	957.97	915.73	
	+Region[1]	894.19	1,214.22	1,089.54	
	+Region[2]	1,124.23	1,042.52	925.80	
	+Region[3]	904.35	1,106.97	911.88	
	+Region[4]	1,086.60	1,031.78	860.41	
	+Category[1]	+All Region[0]	939.83	1,082.63	977.79
	+Category[2]	+All Region[0]	988.25	850.40	959.15

TABLEAU JPivot

			Mesures			
Region	Department	Positions	Actual	Budget	Variance	Variance Percent
-All Regions	+All Departments	+All Positions	143,639,982.00	143,199,389.00	-440,593.00	-31%
Central	+All Departments	+All Positions	37,893,162.00	38,397,600.00	-504,438.00	1.31%
Eastern	+All Departments	+All Positions	35,248,940.00	35,487,861.00	-238,921.00	.67%
Southern	+All Departments	+All Positions	35,248,940.00	34,803,861.00	445,079.00	-1.28%
Western	+All Departments	+All Positions	35,248,940.00	34,510,067.00	738,873.00	-2.14%

Slicer:

Drill Through Table for Actual			
Region	Department	Positions	Actual
Southern	Sales	District Manager	700 000,00
Southern	Sales	Senior Sales Rep	421 200,00
Southern	Sales	Sales Rep	690 000,00
Southern	Sales	Account Executive	290 000,00
Southern	Sales	Pre-Sales	650 000,00
Southern	Executive Management	CEO	500 000,00
Southern	Executive Management	SVP WW Operations	249 800,00
Southern	Executive Management	SVP Strategic Development	226 000,00
Southern	Executive Management	SVP Partnerships	531 780,00
Southern	Finance	CFO	831 800,00

Page 1/4 Aller à la page 1

TABLEAU JPivot

L'utilisateur devra cependant faire attention à ne remonter du datawarehouse qu'un nombre limité données et à retirer de la vue les informations inutiles, non pas pour des questions de performance de Mondrian, mais tout simplement car des graphiques générés sur un trop grand nombre de données sont difficilement lisibles et car les interfaces web deviennent moins efficaces lorsqu'il s'agit de charger des dizaines de pages de données.

Conclusion

JPivot et Mondrian forment un couple performant, capable de traiter des cubes volumineux. On retrouve aujourd'hui ce couple dans toutes les plateformes décisionnelles open source.

FreeAnalysis

FreeAnalysis¹² est un ensemble de logiciels développés autour du serveur multidimensionnel Mondrian. Anciennement nommé FreeOlap, FreeAnalysis regroupe un client OLAP web doté d'une interface AJAX, un client OLAP de type client lourd, ainsi qu'un outil de conception de cubes pour les développeurs.

Client Web : FreeAnalysis Web

Tout comme JPivot, cette application permet de manipuler des cubes Mondrian via une interface web.

FreeAnalysis propose une interface conviviale : les différentes dimensions du cube sont présentées sous forme d'arbre et il est possible de les ajouter au tableau croisé par simple glisser/déposer.

Le reste de l'interface profite également de l'utilisation de l'AJAX en étant très réactive et propose la plupart des fonctions proposées par ce type d'applications : drill down, drill through, rotations, graphiques, exports,

¹² <http://sourceforge.net/projects/freeanalysis/>

The screenshot shows the FreeAnalysis software interface. On the left is a tree view with categories: Region (All Regions, Central, Eastern, Southern, Western), Department (All Departments, Executive Management, Finance, Human Resource, Marketing & Communication, Product Development, Professional Services, Sales), and Positions (All measures, Actual, Budget, Variance). The main window displays a table with the following data:

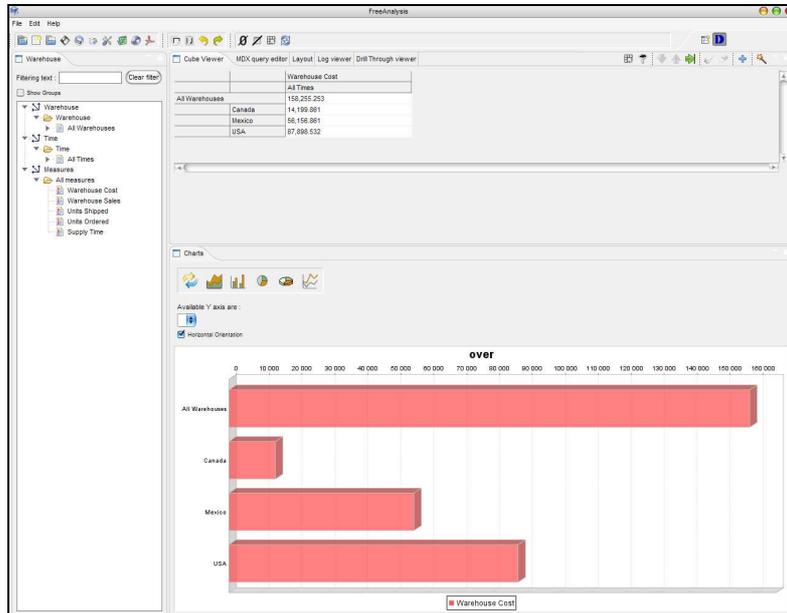
		All Departments	Executive Management	Finance	Human Resource	Marketing & Communication	Product Development	Professional Services
Budget	All Regions	Actual	143,639,862.00	6,299,022.00	12,224,220.00	13,075,483.00	13,910,753.00	10,644,102.00
	Central		37,893,162.00	1,776,282.00	3,106,680.00	3,436,663.00	3,590,423.00	2,997,702.00
	Eastern		35,248,940.00	1,507,580.00	3,039,180.00	3,212,200.00	3,440,110.00	2,548,800.00
	Southern		35,248,940.00	1,507,580.00	3,039,180.00	3,212,200.00	3,440,110.00	2,548,800.00
	Western		35,248,940.00	1,507,580.00	3,039,180.00	3,212,200.00	3,440,110.00	2,548,800.00
Variance	All Regions		143,639,862.00	6,299,022.00	12,224,220.00	13,075,483.00	13,910,753.00	10,644,102.00
	Central		37,893,162.00	1,776,282.00	3,106,680.00	3,436,663.00	3,590,423.00	2,997,702.00
	Eastern		35,248,940.00	1,507,580.00	3,039,180.00	3,212,200.00	3,440,110.00	2,548,800.00
	Southern		35,248,940.00	1,507,580.00	3,039,180.00	3,212,200.00	3,440,110.00	2,548,800.00
	Western		35,248,940.00	1,507,580.00	3,039,180.00	3,212,200.00	3,440,110.00	2,548,800.00

FREEANALYSIS

Client Lourd : FreeAnalysis

On retrouve dans le client lourd de FreeAnalysis la plupart des fonctionnalités proposées par les applications de manipulation des cubes.

FreeAnalysis se connecte en XML/A sur un serveur d'analyse multidimensionnelle, il nécessite par conséquent l'utilisation d'une plateforme décisionnelle centrale dans l'entreprise.



FREEANALYSIS

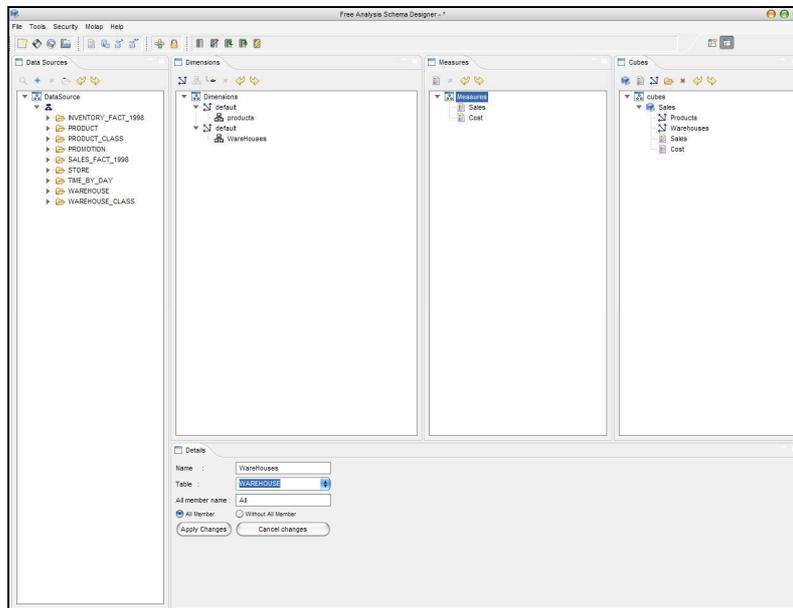
www.smile.fr

Conception de cubes : Schema Designer

FreeAnalysis propose un outil de conception de cubes : « Schema Designer ». Il permet de créer et mettre à jour des cubes Mondrian, sans passer par l'édition de fichiers XML.

La création d'un cube passe par 3 grandes étapes :

- La connexion au datawarehouse et la sélection des tables souhaitées.
- La déclaration des dimensions et des mesures.
- La création de cubes qui utilisent les dimensions et mesures déclarées précédemment.



FREEANALYSIS

Conclusion

FreeAnalysis manque encore de finition. Si les grandes fonctions sont là et fonctionnent parfaitement, il manque encore certaines possibilités et indicateurs visuels nécessaires à l'utilisateur (Drill-replace, icônes dans les tableaux croisés, détails dans les graphiques, désactivation des fonctions non disponibles). Des erreurs surviennent également régulièrement dans les différentes applications.

Dans l'état actuel, **l'utilisation de FreeAnalysis ne sera pas recommandée**, néanmoins ces applications ont un très fort potentiel et sont à surveiller de très près car elles apportent des nouveautés très attendues : AJAX dans les interfaces Web, outil graphique de conception de cubes, ...

Palo

Palo¹³ est un serveur OLAP développé en .NET et distribué sous licence GPL. Il fait partie de la catégorie M-OLAP : **toutes les données sont chargées en mémoire** et calculées à la volée, ce qui lui permet d'obtenir de très bons temps de réponse.

Dans sa version gratuite, il s'intègre **dans le tableur Microsoft Excel** – ou OpenOffice Calc - sous forme de plugin et permet d'insérer des données multidimensionnelles dans les feuilles de calcul. Cette intégration permet de tirer parti des fonctionnalités de tableur et graphiques d'Excel.

www.smile.fr

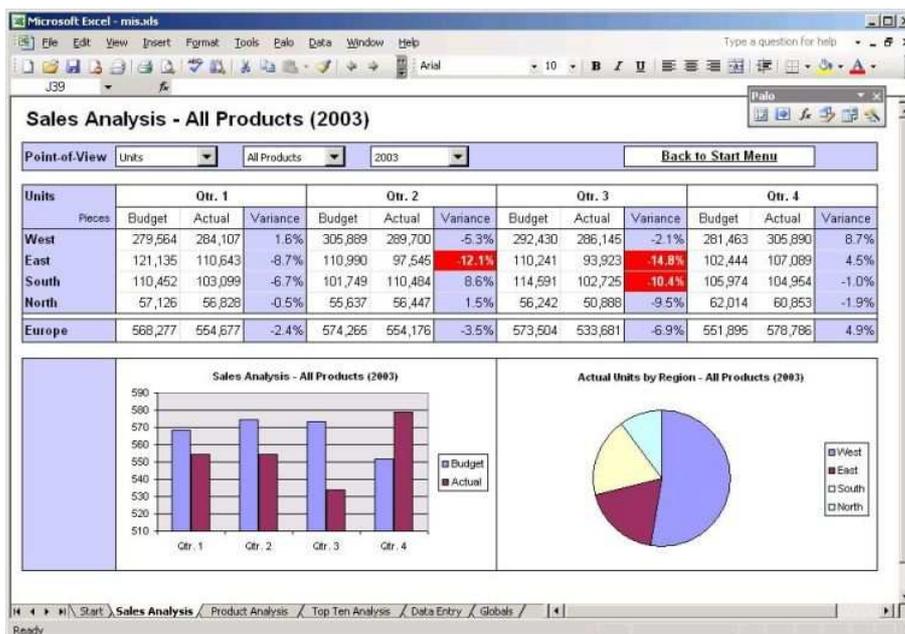


TABLEAU DE BORD DANS MS EXCEL, CONSTITUE AVEC PALO

Palo n'est cependant pas limité à Excel et dispose d'API vers les langages Java et PHP. Il peut ainsi être interrogé depuis d'autres applications, telles que Palo Eclipse Client (JPalo cf. ci-dessous).

Palo ne supporte pas le langage MDX et nécessite l'utilisation d'applications clientes dédiées. Il n'est donc pas possible de l'utiliser conjointement avec JPivot par exemple.

¹³ <http://www.palo.net>

Dans le cadre d'un projet décisionnel, Palo va devoir utiliser les données de l'entreprise. Contrairement à Mondrian qui travaille directement sur une base relationnelle, les données doivent être importées dans la base de données Palo. Des API sont fournies et permettent d'écrire dans le cube M-OLAP depuis les langages C, Java, PHP et .NET.

De plus, un ETL est fourni par Jedox pour alimenter cette base, et les ETL open source Pentaho Data Integration et Talend disposent de connecteurs vers PALO.

On notera qu'il est également possible, depuis Excel, d'importer les données à partir d'un fichier texte ou d'une connexion ODBC mais cette solution n'est évidemment pas adaptée à l'import automatisé d'un grand nombre de données.

Palo propose de plus une fonctionnalité différentiate et très intéressante : **la possibilité de saisir ou modifier les valeurs**. Les valeurs peuvent être gérées selon différentes phases : budget initial, réalisé, budget révisé ..., ce qui permet de construire une application de saisie décentralisée et de consolider le tout in fine, ou à l'inverse de saisir des données globales et de les ventiler ensuite.

Dans sa version payante (et non open source), Palo permet en outre de **basculer un template de saisie ou de restitution Excel en mode web** ! Cette fonctionnalité, gérée par le Worksheet Server, est fondée sur un moteur de conversion Excel/.Net vers PHP, fonctionnalité tout à fait unique sur le marché !

- ❖ Elle présente en lignes/colonnes (comme dans Excel) un template de saisie, sans bien-sûr de possibilité de rotation ou de drill.
- ❖ Elle permet en outre de gérer nativement le mode Read-Only sur certains indicateurs et le mode Read-Write sur d'autres.

Cette fonctionnalité permet donc d'administrer **en mode web un workflow de saisie contrôlé** : extrêmement utile non seulement pour l'élaboration budgétaire, mais aussi pour centraliser de façon contrôlée tout type d'information décisionnel à travers le web.

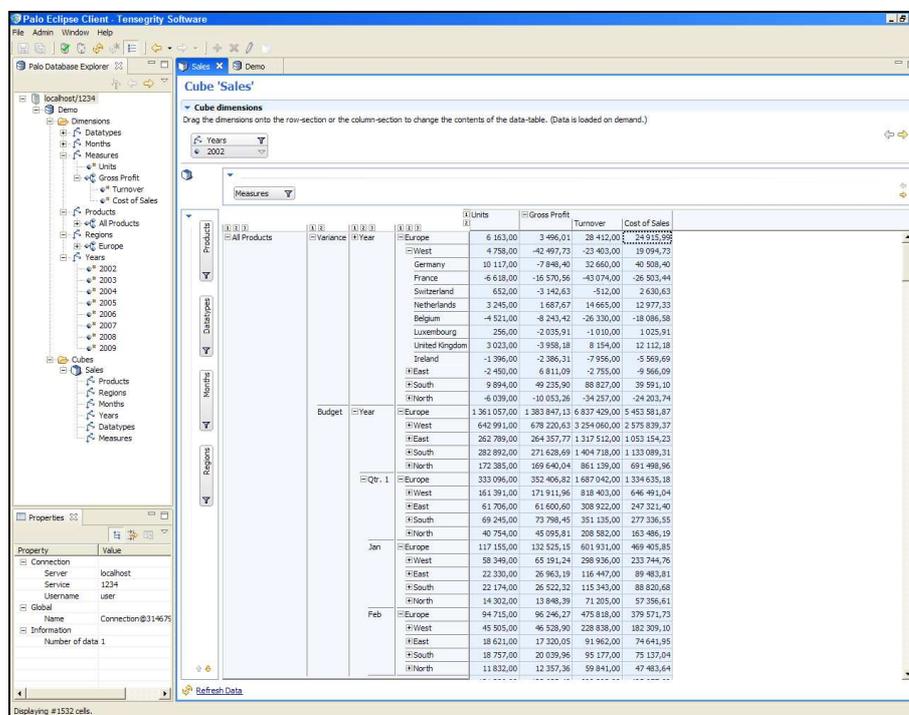
Palo est déjà largement utilisé dans son pays d'origine, l'Allemagne, comme outil de planification et d'analyse. Il n'est pas un outil métier «off the shelf» d'élaboration budgétaire, mais toutes les fonctions sont présentes pour le faire.

JPalo

Le projet JPalo¹⁴ propose des outils pour le serveur Palo développés en Java et notamment le client OLAP « Palo Eclipse Client »

Palo Eclipse Client est une application de type client lourd qui s'intègre à l'environnement de développement Eclipse. Elle permet de présenter les données d'un cube Palo sous forme d'un tableau croisé. L'utilisateur peut ensuite effectuer les opérations classiques de l'analyse multidimensionnelle.

www.smile.fr



PALO ECLIPSE CLIENT

Dans sa version actuelle, Palo Eclipse Client ne propose pas la représentation des données sous forme de graphiques.

JPalo dispose aussi d'une interface web relativement agréable à utiliser, pouvant s'interfacer avec différents serveurs OLAP, dont Mondrian.

¹⁴ <http://www.jpalo.net/>

Cube 'Sales'

Dimensions
Drag dimensions onto the row section or the column-section to change contents of the data-table. (Data is loaded on demand.)

Datatypes: Variance, Measures: Units

Years: 2002, 2003, 2004, 2005, 2006
Months: Qtr. 1, Qtr. 2, Qtr. 3, Qtr. 4

Product	2002				2003	2004	2005	2006	
	Year	Qtr. 1	Qtr. 2	Qtr. 3	Qtr. 4	Year	Year	Year	
All Products	55 648,97	35 016,05	251,94	16 759,98	3 621,00	-47 164,00	98 026,00	-828 230,70	448 721,...
Stationary PC's	2 181,60	3 241,00	-13 021,00	9 034,00	2 927,60	-37 601,00	36 310,00	-333 202,09	151 150,
Portable PC's	13 102,55	7 665,06	2 828,93	3 795,00	-1 186,44	27 777,00	62 058,00	-217 594,96	169 045,
Monitors	38 466,28	23 392,00	11 223,00	2 973,00	878,28	-16 520,00	9 583,00	-239 818,58	120 574,
Peripherals	1 898,54	717,99	-778,98	957,98	1 001,56	-20 820,00	-9 925,00	-37 615,08	7 951,

Regions: [Dropdown]

Refresh

JPALO – INTERFACE WEB

www.smile.fr

Weka

Présentation

Weka¹⁵ est un outil permettant d'exécuter des algorithmes de data-mining sur un ensemble de données. Il est ainsi possible d'isoler des populations ou d'extraire des règles à partir des données contenues dans le datawarehouse.

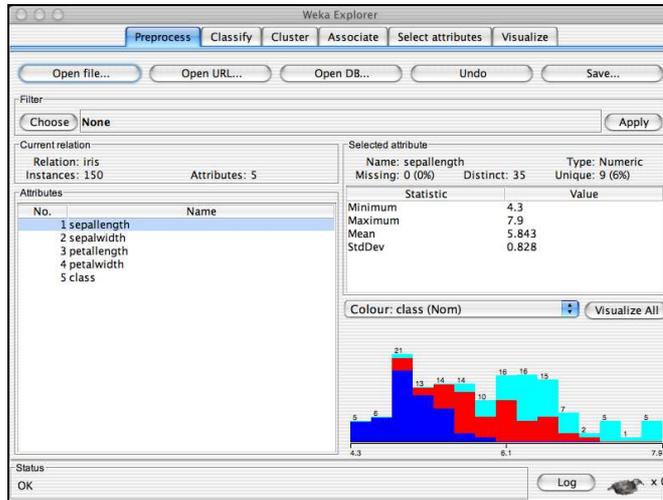
Il se présente sous la forme d'une application indépendante, disposant d'une interface utilisateur graphique ou en ligne de commande.

L'utilisateur peut appliquer un à un les différents algorithmes, ou bien créer, d'une façon similaire à PDI, un workflow de traitements qui pourra par exemple être utilisé dans une plateforme décisionnelle pour analyser périodiquement les données.

L'utilisation de Weka demande de bonnes connaissances du data-mining et des différents algorithmes statistiques utilisés.

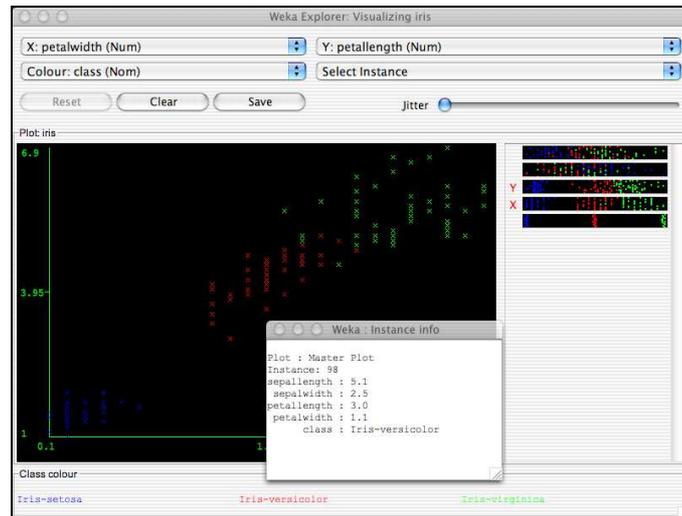
¹⁵ <http://www.cs.waikato.ac.nz/~ml/index.html>

Décisionnel – Solutions open source



WEKA - PRETRAITEMENT

www.smile.fr



WEKA - VISUALISATION DES DONNEES

LES SUITES DECISIONNELLES

Nous avons étudié les principales plateformes décisionnelles disponibles en open source. Ces plateformes utilisent toutes des technologies Java.

Elles partagent certains des composants open source que nous venons de présenter, c'est pourquoi leurs fonctionnalités de base restent assez proches.

La présentation se concentrera sur les spécificités de chaque plateforme et ne détaillera donc pas les possibilités des composants communs.

Pentaho

Présentation

Pentaho¹⁶ est une suite logicielle qui permet la distribution de fonctionnalités et documents décisionnels à un grand nombre de personnes par l'intermédiaire d'une application Web ou un portail.

Au début, Pentaho était proposé à la fois en version communautaire sous licence open source et en version professionnelle sous licence commerciale, avec des modules supplémentaires pour cette dernière licence, comme le portail personnalisé, l'abonnement aux documents et la gestion du cycle de vie des documents.

Depuis 2007 et notamment la version 1.6, **Pentaho a basculé en open source la quasi-intégralité des modules**, actant ainsi son **orientation claire dans le monde open source**. Son modèle économique est aujourd'hui fondé sur la diffusion de la version Enterprise qui concentre toutes les innovations et les correctifs de l'éditeur, version payante, ainsi que sur les services classiques d'un éditeur (formation, support, expertise). Les nouveautés de cette version payante étant ensuite basculées en open source dans leur intégralité au bout de 6 à 12 mois.

¹⁶ <http://www.pentaho.com>

Pentaho propose un plugin pour Microsoft Excel – sous licence commerciale - qui permet d'exploiter les données du datawarehouse dans les tableaux croisés dynamique du tableur. Ce plugin interroge en XML/A les hypercubes de Mondrian, ce qui donne aux utilisateurs une autre manière d'attaquer les données qu'ils peuvent apprécier, même si l'export vers Excel était déjà disponible dans JPivot.

Fonctionnalités

Pentaho permet, via une interface unique, l'accès aux fonctionnalités des différents outils décisionnels open source présentés dans ce document. L'interface peut être soit un portail J2EE (Pentaho utilise le portail JBoss Portal), soit une application Web « classique ».

Il est ainsi possible de consulter des états (les outils BIRT, JasperReports et PentahoReport sont supportés), d'utiliser les fonctions d'exploration de données de Mondrian/JPivot, et de constituer des tableaux de bord.

Lors de l'ouverture d'un document, l'utilisateur a la possibilité de modifier les paramètres d'exécution, par exemple en choisissant la période concernée ou les produits représentés.

La plateforme permet également d'envoyer automatiquement par email un ensemble d'états à leurs destinataires respectifs. Cet envoi se produit soit périodiquement, soit dès qu'une certaine condition est réalisée.

La plateforme Pentaho est organisée autour de deux composants : un moteur de workflow et un planificateur de tâches.

Moteur de workflow

Le moteur de workflow utilisé est « Enhydra Shark »¹⁷. Son rôle est de structurer l'exécution des processus décisionnels :

Chaque fonctionnalité (génération d'états, analyse) est appelée par l'exécution d'une Action, ou « Xaction » dans le langage Pentaho.

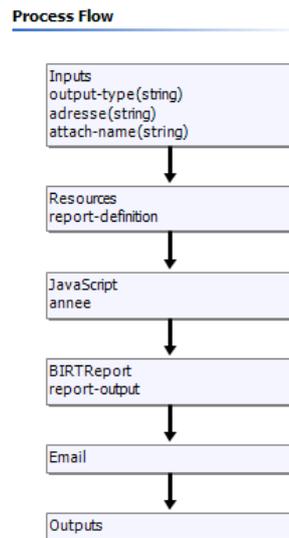
Une «Xaction est une succession d'étapes simples, telles que l'exécution d'une requête sur une base de données, la génération d'un état, l'envoi d'un mail ou l'exécution d'un job de PDI. Les traitements effectués dans ces actions peuvent

17 <http://www.enhydra.org/workflow/shark/index.htm>

Décisionnel – Solutions open source

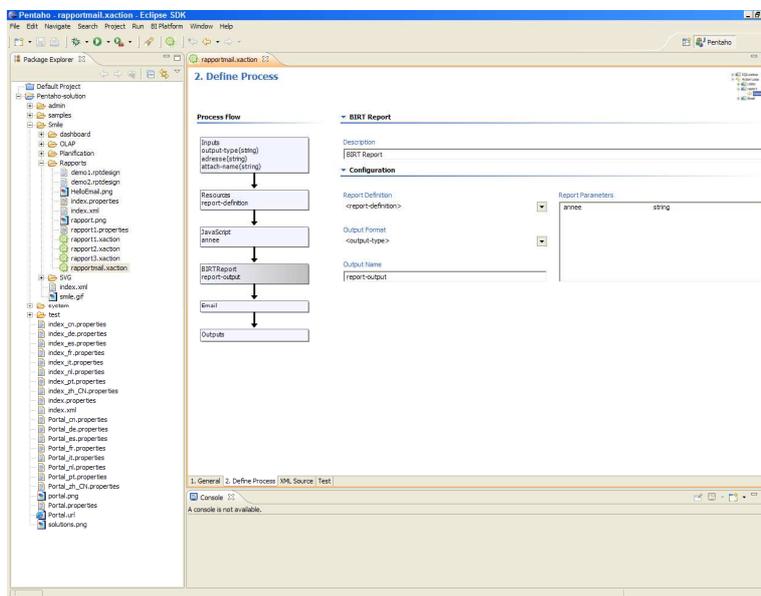
être assez poussés, car il est possible de traiter séparément chaque élément d'un ensemble de résultats ou bien de réaliser des calculs spécifiques via l'exécution d'un programme javascript.

Le moteur de workflow gère l'enchaînement des différentes étapes.



ENVOI D'UN ETAT PAR MAIL

Chaque Xaction est décrite par un simple fichier XML. Pentaho dispose d'un outil, Pentaho Design Studio, qui s'intègre dans l'environnement de développement Eclipse et permet de créer graphiquement ces fichiers.



PENTAHO DESIGN STUDIO

Il faut noter que tous les composants décisionnels supportés par Pentaho constituent une étape du workflow, ce qui permet de combiner leurs fonctionnalités, et d'étendre très facilement les possibilités de la plate forme en développant de nouvelles étapes, pour **en faire une véritable application métier**.

Planificateur

Le planificateur utilisé est « Quartz »¹⁸. Il permet de planifier n'importe quelle action Pentaho.

Il est ainsi possible de créer une action qui génère et envoie par mail un état, et de la planifier pour une exécution périodique.

Le contrôle du planificateur (ajout, suspension, et suppression des actions), peut s'effectuer soit manuellement, soit en utilisant les étapes correspondantes dans une action.

Pentaho permet donc non seulement d'utiliser des outils décisionnels avec une interface unique, mais également de combiner leurs fonctionnalités et de les enrichir grâce à des traitements réalisés en amont et en aval de leur exécution.

www.smile.fr

18 <http://www.opensymphony.com/quartz/>

Interface utilisateur

Il est possible d'utiliser les fonctionnalités de la plateforme Pentaho dans une application Web, un portail J2EE ou par l'appel de web-services.

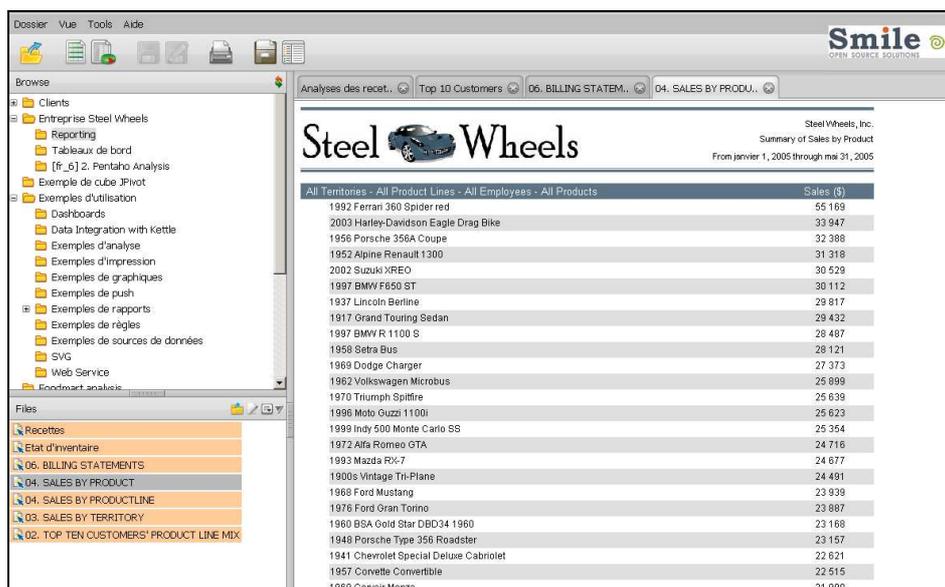
Pentaho inclut deux interfaces pour la présentation et d'exécution des différentes actions disponibles.

◆ Application web

L'interface par défaut de Pentaho, qui donne accès de façon très simple à l'ensemble des actions disponibles ainsi qu'aux fonctions d'administration.

On notera que les comptes utilisateurs sont maintenant supportés, et qu'il ainsi possible limiter l'accès aux différentes actions, mais également de retourner des documents adaptés à chaque profil.

www.smile.fr



APPLICATION WEB PENTAHO 2.0

Cette interface constitue la façon la plus simple de mettre en place la plateforme Pentaho et elle sera donc bien souvent utilisée en priorité.

◆ Portail

Pentaho dispose également de modules adaptés pour un portail J2EE (comme JBoss Portal, supporté par défaut, ou encore Liferay) composé de portlets, sortes de mini-

Décisionnel – Solutions open source

applications, qui donnent accès aux différentes fonctionnalités de la plateforme.

- EmbeddedReportPortlet : permet d'intégrer un état directement dans le portail. Ce portlet peut également être utilisé pour composer des tableaux de bord à base de d'états légers.
- ChartPortlet et WidgetGridPortlet qui permettent de combiner des graphiques afin de créer des tableaux de bord.
- Il existe également un portlet permettant de parcourir l'ensemble des actions disponibles, mais son mode d'affichage n'est pas vraiment adapté aux portails.

www.smile.fr



TABLEAU DE BORD

Regional Report
Region: Central

Department	Position	Actual	Budget	Variance
Executive Management				
CEO		\$549 625	\$522 210	\$27 415
SVP Operations		\$476 000	\$725 887	-\$249 887
SVP Strategic Development		\$183 242	\$403 405	-\$220 163
SVP Partnerships		\$367 415	\$392 100	-\$24 685
Total		\$1 576 282	\$2 043 602	-\$467 320
Finance				
CFO		\$710 272	\$719 855	-\$9 583
Administrative Assistant		\$827 861	\$760 990	\$66 871
IS		\$570 759	\$577 346	-\$6 587
Controller		\$570 979	\$771 070	-\$200 091
Payroll		\$167 415	\$432 100	-\$264 685
Total		\$3 106 880	\$3 067 361	\$39 519
Human Resource				
Administration		\$549 625	\$552 210	-\$2 585
SVP HR		\$574 895	\$570 300	\$4 595
HR Compliance		\$636 199	\$771 253	-\$135 054
HR Training		\$397 473	\$443 570	-\$46 097
Recruitment		\$530 473	\$538 570	-\$8 097
EOE		\$530 207	\$538 380	-\$8 173
Total		\$3 438 863	\$3 414 285	\$24 578

ETAT BIRT DANS PENTAHO

Dans ce mode, la sécurité et la personnalisation de l'interface sont gérés par le portail. Il est par exemple possible de choisir quels portlets seront affichés pour tel ou tel utilisateur. Comme avec l'interface « classique », les informations utilisateur sont propagées jusqu'aux actions Pentaho, ce qui permet, de retourner des documents adaptés, profilés.

Les portlets fournis par Pentaho sont à utiliser lorsqu'on souhaite intégrer la plateforme dans des portails ou intranets d'entreprise existants.

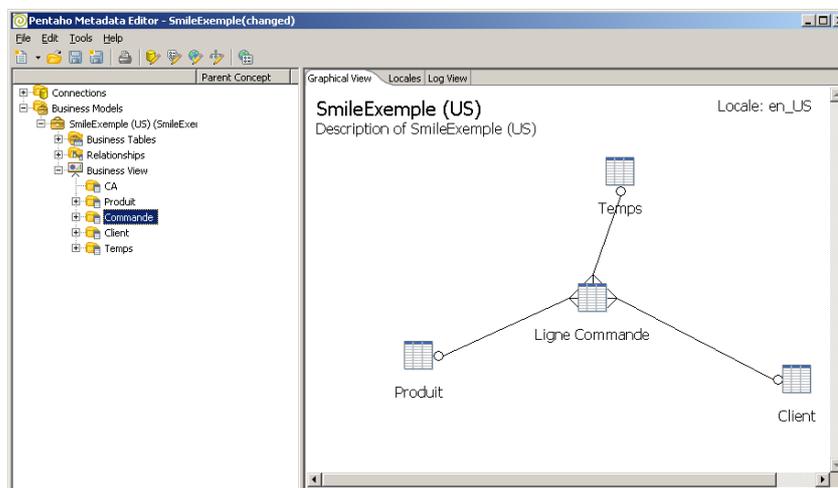
Metadata

Une des évolutions importantes de Pentaho réside dans le **Metadata editor**, qui permet de créer des « vues métier » de bases de données, compréhensibles par les utilisateurs finaux. L'utilisateur ne voit plus directement les tables de la base mais des regroupements d'informations, retranscrites en français (ou en anglais ou en plusieurs langues selon les utilisateurs), potentiellement issus de plusieurs tables et affichés ensemble.

Une fois ces vues métier définies, une interface web « Web Adhoc Query » leur permet de créer des requêtes sans passer par du SQL, en sélectionnant à la souris les champs à afficher, et les regroupements à faire.

www.smile.fr

Le Metadata Editor enregistre le modèle sous le format « Common Warehouse Model », **standard ouvert** de description d'entrepôt.



PENTAHO METADATA EDITOR

Composants utilisés

Pentaho intègre la plupart des composants décisionnels open sources existants.

- Etats : Eclipse BIRT, JasperReports, JFreeReport
- Graphiques : JFreeChart
- Analyse : JPivot, Mondrian
- Portail : JBoss Portal
- Planificateur : Quartz
- Workflow : Enhydra Shark
- ETL : Kettle
- Data-mining : Weka

Bien que trois générateurs d'états différents soient supportés, on choisira en général celui qui correspond le plus aux fonctionnalités souhaitées. L'utilisation cote à cote de deux générateurs peut également s'envisager par exemple dans le cas où l'on souhaite disposer d'un générateur d'états suffisamment intuitif pour être utilisé par les utilisateurs

finaux (BIRT) et de certains rapports utilisant des données issues de requêtes MDX (PentahoReport).

Communauté

Le projet Pentaho est actuellement très actif et démontre l'activité d'un grand éditeur de logiciel.

Plusieurs projets décisionnels open source tels que Kettle, Mondrian, Weka et JFreeReport ont rejoint le projet Pentaho, ce qui garantit une **parfaite intégration** et une **évolution cohérente de l'ensemble**.

La communauté autour du projet est également dynamique comme en témoigne l'activité sur le forum Pentaho ainsi que les diverses contributions faites au projet par les utilisateurs.

Conclusion

Pentaho dans sa version 2.0.1 est une plateforme décisionnelle extrêmement complète et auto-suffisante.

Elle permet non seulement d'utiliser les différents outils décisionnels open source depuis une interface unique, mais elle permet d'étendre et de combiner leurs fonctionnalités grâce à l'utilisation d'un moteur de workflow.

Spago BI

Présentation

SpagoBI¹⁹ est une plate forme décisionnelle développée par la société italienne Engineering Ingegneria Informatica²⁰.

Cette plate-forme est **uniquement distribuée sous licence open source**. Il n'y a donc pas de fonctionnalités absentes et réservées pour une version commerciale comme c'est le cas pour Jasper.

Fonctionnalités

SpagoBI intègre les générateurs d'états Jasper Reports et BIRT pour la génération de rapports simples ou paramétrés.

Il permet également l'analyse de données avec JPivot/Mondrian.

SpagoBI comporte également un **composant d'aide à la création de requêtes** (Query By Example) qui permet à l'utilisateur de sélectionner graphiquement les données à afficher de la base de production (pas obligatoirement d'un entrepôt de données) puis d'effectuer des tris, groupements, sélections et d'exporter le résultat.

L'interface graphique de SpagoBI utilisait le portail eXo platform²¹, ce qui apportait une certaine lourdeur dans l'administration au quotidien, mais la version 2.0 permet de s'en affranchir, même si l'intégration à un portail intranet reste possible.

Un ensemble de portlets sont disponibles pour afficher graphiques, états et indicateurs. Certains indicateurs et graphiques sont capables d'actualiser les données affichées au fur et à mesure, sans intervention de l'utilisateur.

Les données sont stockées dans un référentiel utilisant la norme JCR (Java Content Repository) ce qui le rend librement interchangeable.

¹⁹ <http://spagobi.eng.it>

²⁰ <http://www.eng.it/>

²¹ <http://www.exoplatform.com/>

Interface utilisateur

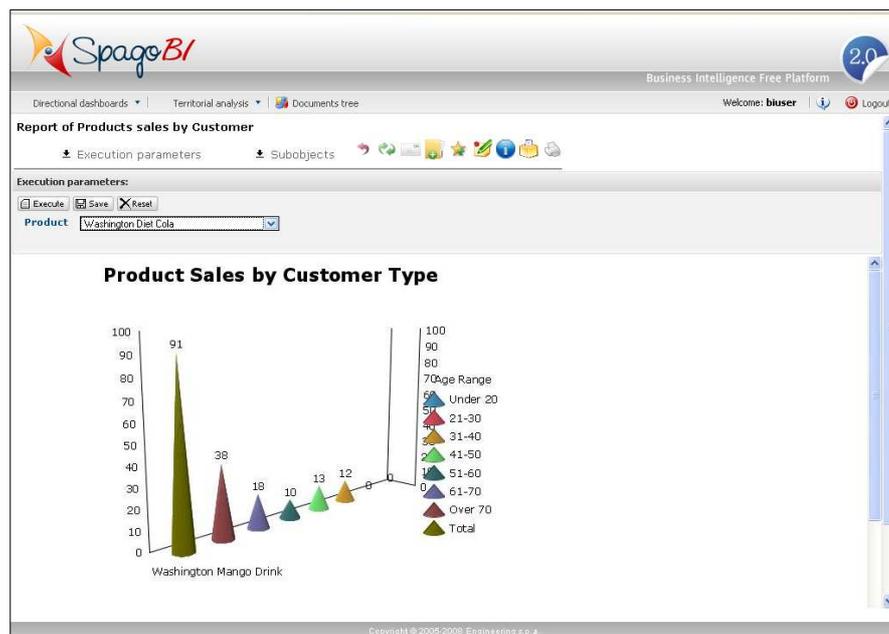
The screenshot shows the SpagoBI interface with a pivot table titled 'Olap Sales over Foodmart'. The table displays sales data categorized by Region, Product, Unit Sales, Store Cost, Store Sales, and Profit. The data is as follows:

Region	Product	Unit Sales	Store Cost	Store Sales	Profit
Central West	Drink	183	140,48	357,58	\$217,10
	Food	1 471	1 230,58	3 095,69	\$1 865,11
	Non-Consumable	373	306,66	776,75	\$470,09
North West	Drink	17 246	13 780,14	34 426,89	\$20 646,75
	Food	134 075	114 371,51	285 597,57	\$171 226,06
	Non-Consumable	35 578	30 557,14	76 270,47	\$45 713,33
South West	Drink	6 688	5 383,08	13 397,54	\$8 014,46
	Food	51 044	43 446,24	108 372,13	\$64 925,89
	Non-Consumable	13 258	11 429,28	28 513,80	\$17 084,52

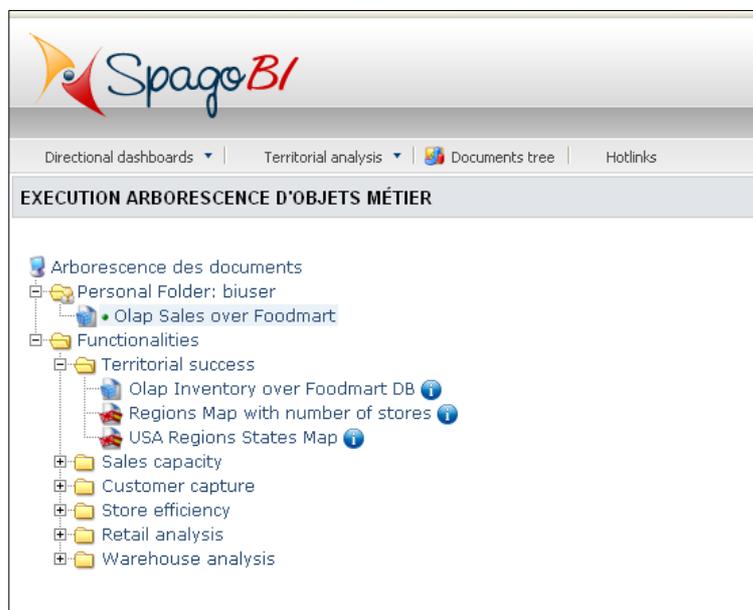
JPIVOT DANS SPAGOBI

La solution prend en charge la sécurité, avec une gestion des rôles utilisateurs pour les pages et groupes de documents décisionnels. Des profils différents sont disponibles pour les utilisateurs, les développeurs et les administrateurs.

www.smile.fr



RAPPORT JASPERREPORT DANS SPAGOBI



NAVIGATION DANS L'ARBORESCENCE DES DOCUMENTS

Workflow de publication

SpagoBI intègre la **gestion du cycle de création, tests et validation des documents décisionnels** (états, cubes multi-dimensionnels,..). Toutes ces opérations sont réalisées directement depuis l'interface d'administration.

Il est possible de créer des états « brouillons » que seuls les développeurs pourront voir et tester, sans qu'ils apparaissent dans l'environnement des utilisateurs.

Les différentes versions de chaque état et des fichiers attachés (gabarits, images,...) **sont enregistrées** et il est possible de revenir à une ancienne version à tout moment.

Cette fonctionnalité permet de mettre à jour les documents décisionnels, sans avoir accès au serveur lui-même et de les tester dans des conditions réelles avant leur mise en production.

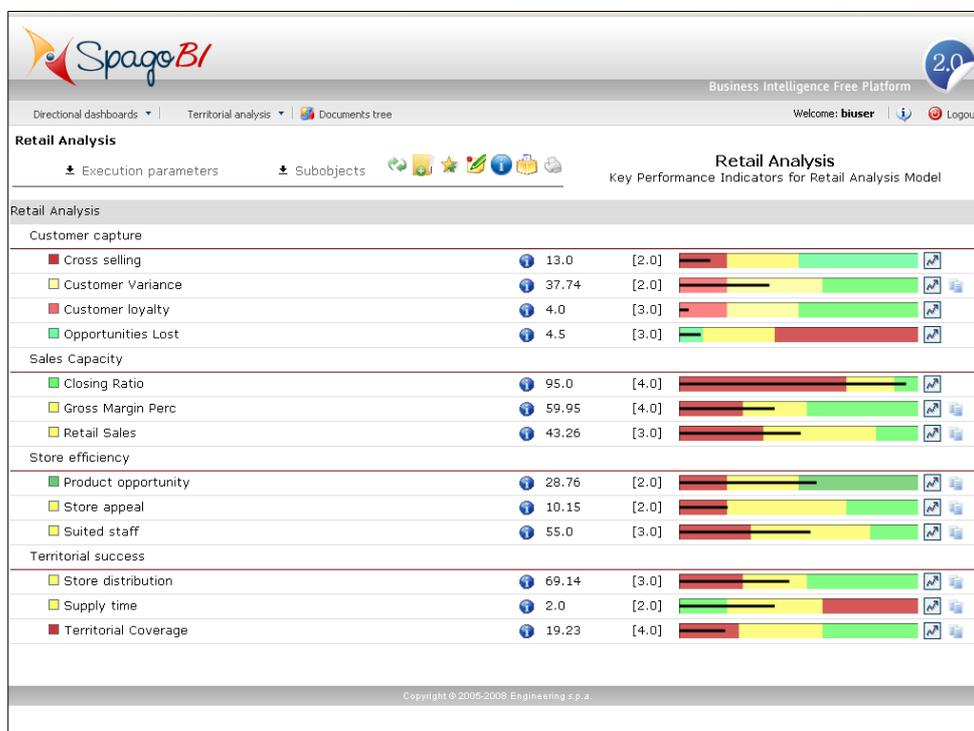
Composants utilisés

SpagoBI utilise plusieurs composants décisionnels open source :

Décisionnel – Solutions open source

- Etats : JasperReports, BIRT
- Analyse : Mondrian, JPivot, JPalo
- Data-mining : Weka
- Portail : eXo platform
- ETL : Talend Open Studio

La fonctionnalité de création de requêtes SQL (Query By Example) est développée spécifiquement par SpagoBI, de même que le moteur d'affichage de KPI



MOTEUR KPI

Intégration des données

SpagoBI a sélectionné Talend comme l'ETL de référence pour cette plateforme. Il est possible de déclencher et planifier des traitements de données directement depuis le portail.

Évolutions

SpagoBI évolue régulièrement : moteur de recherche, workflow, nouveaux moteurs d'analyse, support de Microsoft

Analysis Services et de Talend ont été ajoutés depuis la publication de la première version de ce livre blanc.

On attend particulièrement les livraisons de SpagoBI Studio et SpagoBI Meta, deux modules complémentaires qui permettront respectivement de créer des rapports avec BIRT et la couche sémantique Qbe, et de modéliser cette couche sémantique.

Communauté

La communauté autour du projet SpagoBI reste moins importante que celle de Pentaho, mais son appartenance à OW2²² permet de créer une dynamique et un partage avec d'autres modules open source .

Conclusion

SpagoBI est une plateforme décisionnelle complète, agrégatrice de composants décisionnels soit développés en propre : KPI, Maps, QbE, soit issus d'outils tiers : Mondrian/JPivot, BIRT, JasperReport, Weka, Microsoft SSRS ... Il existe même un connecteur pour Business Objects.

SpagoBI a su dès le début proposer des fonctionnalités de gestion du système décisionnel, avec le versionnement, les workflows, les commentaires, la contextualisation, la gestion des utilisateurs et des droits ... ce qui en fait un outil très intéressant et très pratique en production.

Jasper Soft

Présentation

JasperSoft²³ est la plateforme décisionnelle de JasperSoft, société qui développe également le générateur d'états JasperReport disponible depuis 2001.

Cette plateforme propose des fonctionnalités de reporting et d'analyse et est disponible sous deux licences : GPL et commerciale.

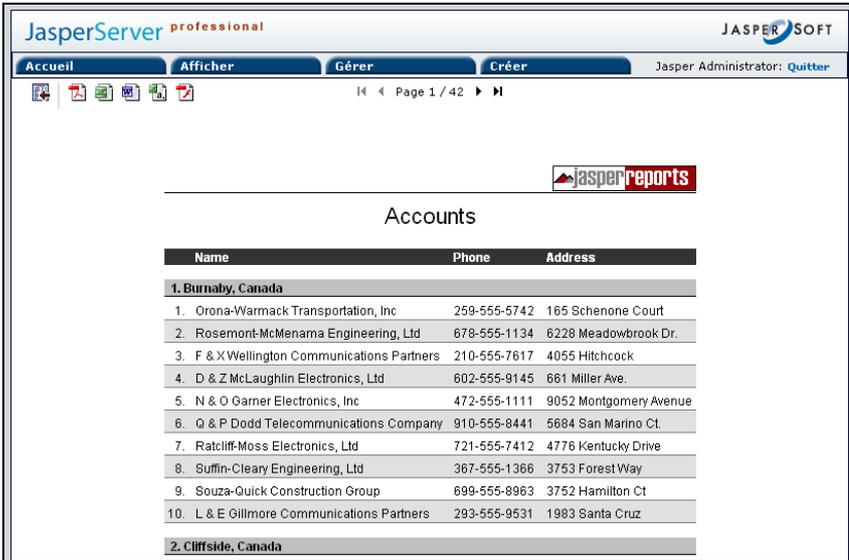
²² <http://www.ow2.org>

²³ <http://www.jaspersoft.com/>

Fonctionnalités

On retrouve dans Jasper toutes les fonctionnalités de JasperReport et JPivot/Mondrian déjà présentées dans ce document. Les utilisateurs ont la possibilité de planifier eux même la génération des rapports et soit de recevoir le document par mail, soit de le stocker dans un répertoire du référentiel.

L'interface n'utilise pas de portail tiers contrairement aux autres plateformes. On dispose ici d'une application web développée spécifiquement qui permet de créer de nouvelles analyses, de publier de nouveaux documents, de gérer les comptes utilisateurs, leurs rôles, ainsi que les droits d'accès aux différents répertoires et fichiers du référentiel.

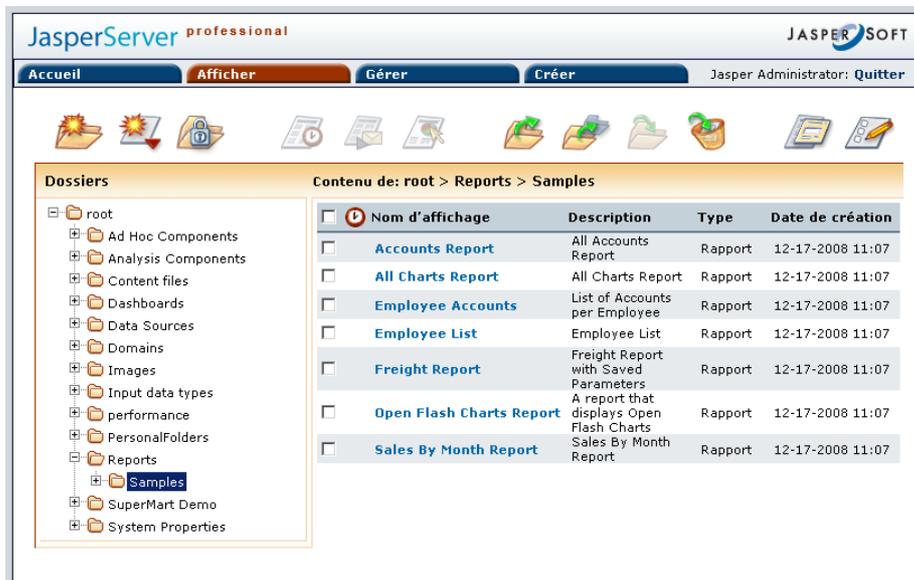


The screenshot shows the JasperServer Professional web interface. At the top, there is a navigation bar with tabs for 'Accueil', 'Afficher', 'Gérer', and 'Créer'. The user is identified as 'Jasper Administrator: Quitter'. The main content area displays a report titled 'Accounts' with a 'jasperreports' logo. The report contains a table with the following data:

Name	Phone	Address
1. Burnaby, Canada		
1. Orona-Warnack Transportation, Inc	259-555-5742	165 Schenone Court
2. Rosemont-McMenama Engineering, Ltd	678-555-1134	6228 Meadowbrook Dr.
3. F & X Wellington Communications Partners	210-555-7617	4055 Hitchcock
4. D & Z McLaughlin Electronics, Ltd	602-555-9145	661 Miller Ave.
5. N & O Garner Electronics, Inc	472-555-1111	9052 Montgomery Avenue
6. Q & P Dodd Telecommunications Company	910-555-8441	5684 San Marino Ct.
7. Ratcliff-Moss Electronics, Ltd	721-555-7412	4776 Kentucky Drive
8. Suffin-Cleary Engineering, Ltd	367-555-1366	3753 Forest Way
9. Souza-Quick Construction Group	699-555-8963	3752 Hamilton Ct
10. L & E Gillmore Communications Partners	293-555-9531	1983 Santa Cruz
2. Cliffside, Canada		

VISUALISATION D'UN RAPPORT DANS JASPERSOFT

A noter : pour les administrateurs, la publication et la modification des différents rapports est possible directement depuis iReport, l'application de conception d'états de JasperReport.



VISUALISATION DU REFERENTIEL DECISIONNEL DANS JASPERSOFT

Composants utilisés

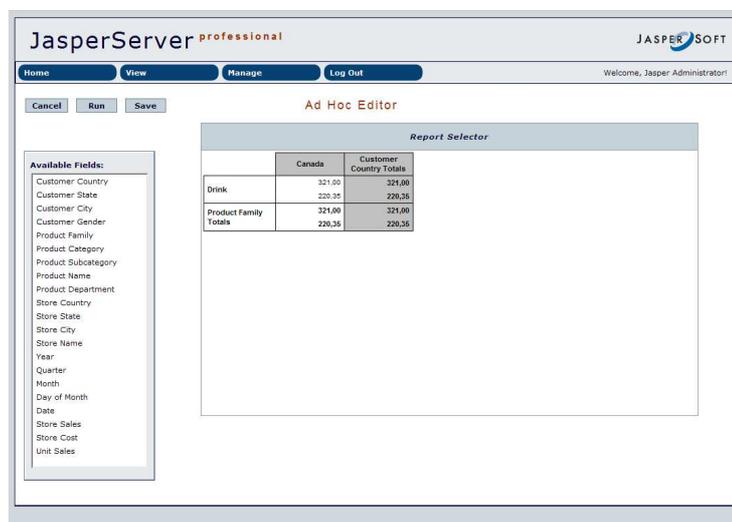
- Etats : JasperReports
- Analyse : Mondrian et JPivot, sous le nom de JasperAnalysis
- Planificateur : Quartz
- ETL : Talend Open Studio, sous le nom de Jasper ETL.

Il est peu probable que cette plateforme supporte d'autres générateurs d'états dans le futur. En effet, JasperSoft a tout intérêt à mettre en avant son propre outil de reporting.

Version professionnelle

La version professionnelle de Jasper Server comporte des fonctionnalités supplémentaires par rapport à la version open source. Les différences majeures sont :

- Un outil de création de rapports (listes, graphiques ou tableaux croisés) en ligne est accessible à tout utilisateur. Il a ainsi la possibilité de créer de nouveaux documents en quelques minutes grâce à une interface à base d'assistants et de glisser/déposer.



CREATION DE TABLEUX CROISES (PRO)

- Un outil de création de dashboards en ligne, accessible à tout profil d'utilisateur.

Les informations mises à disposition pour créer ces rapports sont définies dans des « domaines », couche sémantique au dessus de la base de données, permettant à la fois de définir un lexique métier et de s'abstraire de la technique et du SQL.

On notera qu'il est possible ici pour l'utilisateur de créer des tableaux croisés, sans passer par la technologie Mondrian, ni par la constitution d'une base en étoile.

De plus, l'utilisateur est libre de rajouter des champs calculés, de changer les formats, de changer les libellés, ...

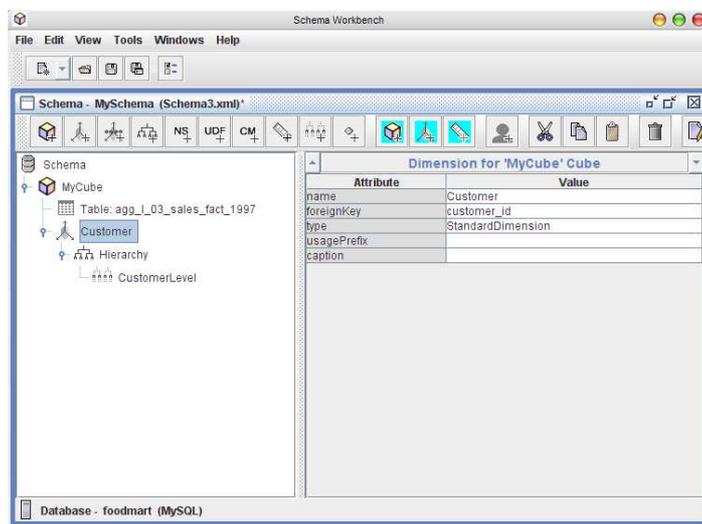
L'interface web de création de rapports est un vrai plus par rapport à la version open source, avec la possibilité, pour les utilisateurs finaux, de construire leur analyse à partir des données mises à disposition, **en toute autonomie**, sans aucune connaissance informatique.

Jasper WorkBench

Jaspersoft fournit une application de conception de cubes Mondrian. L'application est une simple surcouche par rapport à l'édition manuelle des fichiers XML : chaque balise est modifiable à l'aide d'un formulaire basique.

Cette application n'a pas de très forte valeur ajoutée, mais fera gagner un peu de temps aux développeurs, en leur évitant les

erreurs de balises XML et en leur proposant au bon moment les noms des tables de la base de données et leurs colonnes associées.



SCHEMA WORKBENCH

Conclusion

JasperSoft n'a pas développé d'ETL, distribuant simplement celui de Talend, ni de solution de datamining. En revanche, Jasper s'est concentré sur l'utilisabilité de sa solution pour les utilisateurs finaux. Le résultat est excellent, avec une ergonomie 2.0, et, pour des utilisateurs non techniques, une très grande autonomie à la création de nouveaux rapports et de nouveaux tableaux de bord.

Les « plus » de Jasper résident principalement dans les fonctionnalités réservées à la version professionnelle. Parce que cette version est soumise à une souscription annuelle, cela va limiter le déploiement au plus grand nombre comme le permet la viralité de l'open source.

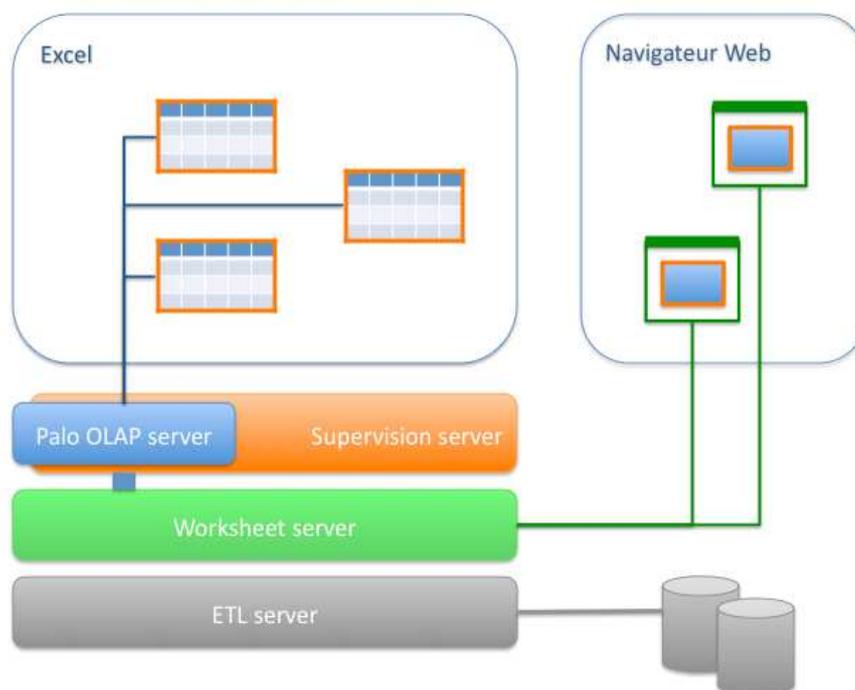
La version professionnelle reste cependant très largement accessible financièrement et permet de bénéficier du support de l'éditeur.

Jaspersoft dispose d'un bureau en France depuis mi 2008.

Jedox

Jedox, société allemande, édite principalement la solution PALO, ainsi que des modules complémentaires, tels que le

- « supervision server » permettant de suivre les modifications faites par les utilisateurs dans les cubes Excel,
- « worksheet server », permettant de diffuser les cubes sur un serveur web, sans client Excel
- « ETL server », ETL de Jedox, communiquant nativement avec la base PALO



Les outils “Palo OLAP Server” et “ETL server” sont open source et libres d’utilisation, avec une seule version – pas de version community ici.

Les outils “supervision server” et “worksheet server” ne sont pas open source et sont soumis à un coût d’acquisition, et à un coût annuel de maintenance comme n’importe quel logiciel propriétaire.

Jedox dispose depuis 2008 d’un bureau en France, ce qui devrait amplifier son déploiement dans l’hexagone.

Décisionnel – Solutions open source

La couverture fonctionnelle de Palo, et les choix technologiques associés, sont très différentiels et ne répondent pas aux mêmes attentes.

On peut supposer de ce fait que Palo convaincra une nouvelle typologie d'utilisateurs, notamment les contrôleurs de gestion, grands concentrateurs d'information et grands consommateurs d'Excel...

SERVICES PROPOSES PAR LES EDITEURS

Si le développement au sein d'une communauté d'un outil technique spécifique a un sens, chacun pouvant utiliser l'outil en question et y apporter ses contributions, le développement d'une suite décisionnelle par une communauté n'existe pas ou pas encore, tant l'effort est important.

Ce sont bien de véritables sociétés qui sont à la tête des différentes suites décisionnelles open source exposés dans ce document.

Aussi, ces sociétés, outre le fait de se partager - ou de se batailler - les composants développés par certaines communautés, essaient de trouver le meilleur moyen de gagner de l'argent dans ce nouveau métier d'éditeur de solutions open source.

Nous constatons deux grandes philosophies :

- Mettre à disposition une version libre de droit avec des fonctionnalités restreintes et faire payer des fonctions complémentaires
 - JasperSoft et Talend ont adopté ce modèle.
 - A noter que les fonctions proposées en mode payant sont présentées comme de confort, mais sont rapidement indispensables pour un déploiement pour toute une entreprise.
- Autoriser la libre utilisation de la solution et offrir des services de support
 - Pentaho et SpagoBI ont choisi ce modèle.
 - Pentaho différencie cependant sa version 'Enterprise', soumise à souscription, garantie par l'éditeur, de sa version 'Community', sans garantie ni support et qui comprend des fonctionnalités et des correctifs de la version Entreprise diffusés en open source avec plusieurs mois de retard.
 - Dans ce cas, avec l'aide d'une bonne équipe informatique en interne, vous pouvez télécharger le

Décisionnel – Solutions open source

produit et réaliser votre projet sans payer quoi que ce soit aux éditeurs concernés. Il vous faudra certes passer un peu de temps en autoformation et assumer les éventuels bugs rencontrés, sans pouvoir espérer d'aide de l'éditeur.

Dans tous les cas, le support proposé par l'éditeur est une garantie qu'en cas de problème, vous ne serez pas seul, et disposerez de l'expertise de l'éditeur pour ne pas rester bloqués tant en mode projet qu'en mode exploitation.

Les contrats de support incluent en général une clause de « Intellectual Property protection », une assurance juridique qui protège le client d'éventuelles actions d'éventuels détenteurs de brevets. Une clause très prisée aux Etats-Unis.

A noter, l'offre de support de Pentaho comprend aussi un support au projet, donnant des réponses sur des problématiques de mises en œuvre, au delà des potentielles anomalies constatées. De plus, Pentaho propose, avec sa version Enterprise, l'accès à de la documentation réservée, et à une console d'administration complémentaire.

CONCLUSION

Les plateformes décisionnelles open source sont aujourd'hui des alternatives complètes et crédibles face aux solutions décisionnelles commerciales. Les différents composants utilisés existent depuis plusieurs années et sont très fiables et éprouvés.

Le Web est le support de prédilection des plateformes open source, qui proposent visualisation, création de rapports et analyse des données sous cette forme.

Si des solutions telles que Pentaho, Jasper ou SpagoBI permettent de mettre en place un portail décisionnel complet, il est également possible d'ajouter facilement des fonctionnalités décisionnelles à une application existante, en utilisant séparément les composants sur lesquels sont basées ces plateformes.

Depuis la première version de ce livre blanc, **les outils évoluent en continu** et, après avoir corrigé leurs défauts de jeunesse, intègrent maintenant des fonctionnalités interactives puissantes, appréciées des utilisateurs et des administrateurs.

De nombreuses entreprises font maintenant le choix des outils décisionnels open source, profitant de leurs performances, de leur capacité d'adaptation à l'existant et de leur faible coût de possession.