

2013

Instituto Politécnico de Coimbra

INSTITUTO SUPERIOR DE ENGENHARIA DE COIMBRA

Análise e Implementação de Open Source Business Intelligence

MESTRADO EM INFORMÁTICA E SISTEMAS

AUTOR | António José Bento Marinheiro

ORIENTADORES | Professor Doutor Jorge Bernardino

Coimbra, dezembro 2013



Instituto Politécnico de Coimbra

Instituto Superior de Engenharia de Coimbra

Departamento de Engenharia Informática e de Sistemas

Análise e implementação de open source Business Intelligence

António Marinheiro

Mestrado em Informática e Sistemas

Coimbra, Dezembro, 2013



Instituto Politécnico de Coimbra

Instituto Superior de Engenharia de Coimbra

Departamento de Engenharia Informática e de Sistemas

Mestrado em Informática e Sistemas

Projecto Industrial

Relatório Final

Análise e implementação de open source Business Intelligence

António Marinheiro

Orientador:

Prof. Doutor Jorge Bernardino

ISEC - DEIS

Coimbra, Dezembro, 2013

Agradecimentos

Concluir esta dissertação é o culminar de um objectivo que se tornou num verdadeiro desafio, nem sempre fácil de manter nos níveis mais elevados de disponibilidade e motivação. Tal só foi possível graças a um conjunto de pessoas que ao longo deste período me apoiaram, aturaram e motivaram a continuar. Deixo os meus agradecimentos:

Ao Professor Doutor Jorge Bernardino que orientou, motivou e acompanhou de forma assídua o meu trabalho, tornando possível esta dissertação.

À minha família por me terem dado a motivação e apoio necessários, por terem aturado o meu nem sempre bom humor, e as minhas ausências em determinado período.

A todos os professores e colegas do Mestrado que directa ou indirectamente contribuíram para a minha formação.

À administração da empresa onde trabalho, e a todos os meus colegas, por me terem apoiado e facultado e tudo quanto necessitei ao longo deste tempo.

Ao Eng. Mário Leitão e ao Eng. Eduardo Jesus por terem facilitado, ajudado e colaborado no processo de implementação em ambiente real, na empresa MICROplásticos, SA.

Por fim, a todos os meus amigos que sempre me ajudaram a superar as maiores dificuldades, prestando, como habitualmente, um apoio essencial. Thanks.

“Emancipate yourselves from mental slavery; None but ourselves can free our minds.”

Bob Marley, in Redemption Song

RESUMO

Vivendo-se actualmente um ambiente de alguma instabilidade no mercado empresarial um pouco por todo o mundo, todo o conhecimento que as empresas adquiram sobre o seu negócio pode traduzir-se numa vantagem competitiva face à concorrência. Nestes momentos de instabilidade é necessário actuar cuidadosamente com um maior conhecimento, havendo uma menor margem de erro nas decisões estratégicas a tomar. Disponibilizar atempadamente o maior número de informações e conhecimento, está na base da escolha das alternativas que se levantam como suporte à decisão, contribuindo assim para um incremento do valor do negócio. Nos últimos anos, as aplicações tecnológicas em diferentes áreas, especialmente de Business Intelligence (BI) desenvolveram-se rapidamente e são consideradas como sendo das de uso mais significativo nas tecnologias da informação. Business Intelligence é uma categoria ampla de aplicações e tecnologias para recolha, armazenamento, análise e disponibilização de acesso a dados para ajudar os utilizadores nas empresas a tomarem melhores decisões de negócio. Considerando que, no passado, o mercado de Business Intelligence foi estritamente dominado por ferramentas comerciais, os tempos recentes foram caracterizados pelo nascimento de soluções open source. Isto representa uma enorme vantagem competitiva, no entanto, a escolha de uma Suite open source BI é um grande desafio. O nosso trabalho contempla o estudo, avaliação e comparação das últimas versões das suites open source de Business Intelligence actualmente existentes. Por fim, iremos seleccionar uma das suites, e realizar uma implementação de uma solução de BI numa empresa em ambiente real, provando que as suites open source BI são uma alternativa a soluções comerciais para a utilização em PMEs. Em suma, pretendemos com este trabalho levar o BI para as empresas, sem custo.

ABSTRACT

Currently living an environment of some instability in the business market all over the world, all the knowledge that companies may acquire about their own business can be reflected as a competitive advantage against the competition. In these moments of instability, it's necessary to act cautiously with greater knowledge, once there's a lower margin for errors in the strategic decisions to make. Timely providing the greatest amount of information and knowledge is the basis for selecting the alternatives that arise as decision support, thus contributing to an increase in the business value. In the latest years, the technological applications in different fields, particularly of Business Intelligence (BI), had a rapid development and are considered to be of the most significant use in information technology. Business Intelligence is a broad category of applications and technologies for gathering, storing, analysing and providing access to data to help the users in the organizations to make better business decisions. Whereas in the past the Business Intelligence market was strictly dominated by closed source and commercial tools, the last few years were characterized by the appearance of open source solutions. This represents a tremendous competitive advantage. However the choice of a suitable open source BI suite is a challenge. The present study evaluated and compared the last versions of the existent open source Business Intelligence suites. Finally, we will select one of the suites and carry an implementation of a BI solution in a real business environment, proving that open source BI suites are a viable alternative to commercial solutions for use in SME. Resuming, with this study we intend to bring BI to the organizations, without costs.

Palavras-Chave

Open Source
Business Intelligence
Suites Business Intelligence
PME

Keywords

Open Source
Business Intelligence
Suites Business Intelligence
SME

ABREVIATURAS

AT&T - American Telephone and Telegraph
BASH - Bourne Again Shell
BDS - Berkeley Software Distribution
BI – Business Intelligence
BIRT – Business Intelligence and Reporting Tools
BRR - Business Readiness Rating
DBMS - Database management system
DFSG - Debian Free Software Guidelines
DSA – Data Staging Area
DW – Data Warehouse
ERP - Enterprise resource planning
ETL – Extract, Transform, Load
FLOSS - Free/Libre as Open Source Software
FSE – Fornecimentos e serviços externos
FSF - Free Software Foundation
GEO – Geocaching
GIS - Geographic information system
GNU - GNU is Not Unix
GPL - GNU General Public Licence
ISO - International Organization for Standardization
KPI - Key performance indicator
LDAP - Lightweight Directory Access Protocol
LGPL - GNU Lesser General Public Licence
MDX - Multidimensional Expressions
MIT - Massachusetts Institute of Technology
MP – Matéria prima
Multics - Multiplexed Information and Computing Service
ODS - Operational Data Store
OLAP - On-line Analytical Processing
OS – Operating System
OSI – Open Source Initiative
PA – Produto acabado
PME – Pequenas e Médias Empresas
QSOS - Qualification and Selection Open Source
ROI - Return on investment
ROLAP - Relational On Line Analytical Processing

SGBD - Sistema de Gestão e Base de Dados

SGDB – Sistema de gestão de base de dados

SIG – Systema de Informação Geográfica

SNC - Sistema de normalização contabilística

SQL - Structured Query Language

TCP/IP - Transmission Control Protocol/Internet Protocol

TI - Tecnologia da Informação

Unics - Uniplexed Information and Computing Service

VAB – Valor acrescentado bruto

XML - Extensible Markup Language

ÍNDICE

CAPÍTULO 1. INTRODUÇÃO	1
1.1. Principais contribuições deste trabalho	2
1.2. Estrutura do relatório.....	3
CAPÍTULO 2. SOFTWARE OPEN SOURCE	5
2.1. A história do software open source	5
2.2. Free Software vs Open Source	7
2.3. Categorias de licenças de software	9
2.4. O open source nas empresas	10
2.5. Exemplos de software open source	13
CAPÍTULO 3. BUSINESS INTELLIGENCE	15
3.1. Dados vs Informação vs Conhecimento vs Decisão	16
3.2. Componentes do BI	17
3.2.1. Dados Operacionais	18
3.2.2. ODS – Operational Data Store	19
3.2.3. Ferramentas ETL	19
3.2.4. Data warehouse e o modelo em estrela	21
3.2.5. Data mining	23
3.2.6. Visualização dos resultados	24
3.3. Benefícios na utilização de BI	25
3.4. Dificuldades na implementação de BI	27
CAPÍTULO 4. SUITES OPEN SOURCE BI.....	29
4.1. Suites estudadas	29
4.1.1. JasperSoft	29
4.1.2. Palo	31
4.1.3. Pentaho	33
4.1.4. SpagoBI	35
4.1.5. Vanilla	37
4.2. Metodologias de avaliação	39
4.2.1. Integração	39

4.2.2. Entrega da informação	39
4.2.3. Análise	40
4.3. Comparação das Suites	41
CAPÍTULO 5. AVALIAÇÃO EXPERIMENTAL	45
5.1. OpenBRR	45
5.2. Avaliação	50
CAPÍTULO 6. MODELO DE DADOS	57
6.1. A empresa	57
6.2. Questões a responder	58
6.2.1. Recursos Humanos	58
6.2.2. Administração	59
6.2.3. Comercial	59
6.2.4. Logística	60
6.3. Modelo dimensional	60
CAPÍTULO 7. IMPLEMENTAÇÃO	67
7.1. Instalação da SpagoBI	67
7.2. MySQL	68
7.3. Talend	69
7.4. Extração da informação	76
7.5. Acessos	83
7.4. Automatização	84
CAPÍTULO 8. CONCLUSÕES E TRABALHO FUTURO	85
ANEXO A - Avaliação das métricas de cada categoria, do método Open BRR	95
ANEXO B - Instalação da suite open source SpagoBI, em Linux	105
ANEXO C - Artigo CISTI'2013 "Analysis of Open Source Business Intelligence Suites"	109
ANEXO D - Artigo C3S2E13 "OpenBRR Evaluation of an Open Source BI Suite" .	120

ÍNDICE DE FIGURAS

Figura 2.1. Logótipo da OSI.....	5
Figura 2.2. Logótipo do GNU	6
Figura 2.3. Logótipo do Copyright e do Copyleft	9
Figura 3.1. Dos dados à decisão	17
Figura 3.2. Arquitectura típica do BI.....	18
Figura 3.3. Modelo em estrela	22
Figura 3.4. Processo da extracção do conhecimento	24
Figura 3.5. Exemplos de gráficos de extracção de informação	25
Figura 4.1. Logótipo da Jaspersoft	29
Figura 4.2. Um aspecto do iReport Designer, da suite Jaspersoft	31
Figura 4.3. Logótipo da Palo	31
Figura 4.4. Um aspecto da suite Palo	33
Figura 4.5. Logótipo da Pentaho	33
Figura 4.6. Um aspecto da suite Pentaho	35
Figura 4.7. Logótipo da SpagoBI	36
Figura 4.8. Um aspecto da suite SpagoBI	37
Figura 4.9. Logótipo da Vanilla	37
Figura 4.10. Um aspecto da suite Vanilla	39
Figura 5.1. Esquema do modelo OperBRR	47
Figura 6.1. Organograma da empresa	58
Figura 6.2. Visão geral do modelo dimensional	63
Figura 6.3. Detalhe do modelo dimensional para a área dos Recursos Humanos	64
Figura 6.4. Detalhe do modelo dimensional para a área da Administração	64
Figura 6.5. Detalhe do modelo dimensional para a área Comercial	65
Figura 6.6. Detalhe do modelo dimensional para a área da Logística	66
Figura 7.1. SpagoBI no browser, preparado para fazer login	67
Figura 7.2. Modelo entidade-relacionamento no MySQL Workbench	68
Figura 7.3. Data warehouse gerada através do MySQL Workbench	69
Figura 7.4. Ligações e queries no repositório do Talend	70
Figura 7.5. Configuração da ligação à base de dados operacional, Oracle	71
Figura 7.6. Job principal no Talend, contendo a sequência de execução dos sub-jobs	72

Figura 7.7. Processo de ETL para a tabela de factos Absentismo	72
Figura 7.8. Configuração do componente QryInputAbsentismo	73
Figura 7.9. Configuração do componente AggregateFaltas	73
Figura 7.10. Configuração do componente de lookup à tabela Data	74
Figura 7.11. Configuração do componente de lookup para os dados do funcionário	74
Figura 7.12. Configuração mapeamento para futura gravação de dados	75
Figura 7.13. Configuração do componente responsável por gravar os dados	75
Figura 7.14. Visualização dos dados inseridos no MySQL Workbench	76
Figura 7.15. Configuração de um documento OLAP no SpagoBI Studio	77
Figura 7.16. Hierarquia da pesquisa OLAP na Area Funcional	78
Figura 7.17. Upload do template OLAP para o servidor SpagoBI	78
Figura 7.18. Explorar o documento AbsentismoArea no SpagoBI	79
Figura 7.19. Formato do ficheiro SVG do mapa do mundo	80
Figura 7.20. Criação do documento “Vendas por Pais 2013” no SpagoBI	81
Figura 7.21. Resultado final do documento “Vendas por Pais 2013”	82
Figura 7.22. Gráfico do peso do Custo do Pessoal relativamente às Vendas	82
Figura 7.23. Menu da gestão de utilizadores no SpagoBI	83
Figura 7.24. Criação do documento de carregamento do ETL no SpagoBI	84

ÍNDICE DE TABELAS

Tabela 2.1 – Exemplos de software open souce em diversos domínios	13
Tabela 4.1 – Comparação das suites open source BI	41
Tabela 5.1 – Categorias da metodologia OpenBRR	45
Tabela 5.2 – Métricas para cada categoria da metodologia OpenBRR	47
Tabela 5.3 – Peso das diversas categorias escolhidas	51
Tabela 5.4 – Peso das capacidades da categoria funcionalidades	52
Tabela 5.5 – Avaliação das funcionalidades das suites	52
Tabela 5.6 – Avaliação da categoria Usabilidade	53
Tabela 5.7 – Avaliação da categoria Qualidade	54
Tabela 5.8 – Avaliação da categoria Documentação	54
Tabela 5.9 – Avaliação da categoria Comunidade	55
Tabela 5.10 – Avaliação da categoria Segurança	55
Tabela 5.11 – Avaliação da categoria Arquitectura	56
Tabela 5.12 – Nota ponderada das suites em avaliação	56

1. INTRODUÇÃO

No ambiente empresarial actual, em virtude das grandes alterações que se têm verificado, as organizações necessitam ser suficientemente ágeis para antecipar as mudanças, encontrando soluções e adaptando-se, ou criando condições para serem mais competitivas relativamente à concorrência.

Neste contexto, há cada vez mais a necessidade das organizações terem informações relevantes que auxiliem no suporte de decisões estratégicas e de ferramentas para servir esses propósitos. Desde os primórdios da estratégia organizacional, na *A Arte da Guerra*, escrito há cerca de 2500 anos por Sun Tzu (Tzu, 2011), que existe uma preocupação na forma de se conhecerem os pontos fortes e fracos das organizações, assim como os dos concorrentes. Mais recentemente, Michael Porter refere que em qualquer empresa a informação tecnológica tem um efeito poderoso na vantagem competitiva em termos de custos e diferenciação (Porter et al., 1985). Antes da tratamento informático da informação, não sendo esta estruturada, as decisões tomadas tinham sempre um grande grau de intuição. Com a complexidade tecnológica e concorrencial dos dias de hoje, as empresas estão obrigadas a prestar uma atenção redobrada ao processo de controlo interno, de desempenho e aos sistemas de suporte à decisão. Neste âmbito, (Williams et al., 2003) afirma que a engenharia de processos e a análise da mudança estabelecem as bases para incrementar o valor do negócio.

As pequenas e médias empresas (PME), representam mais de 98% das empresas de todo o mundo (Global Alliance of SMEs, 2010). No entanto a sua capacidade financeira é normalmente reduzida, o que limita muitas vezes o acesso a sistemas que ajudem na tomada de decisão, bem como o próprio acompanhamento do seu negócio, o que se não for feito adequadamente conduz a uma perda da competitividade da empresa, ou até à sua ruína.

Nos últimos anos, as aplicações tecnológicas em diferentes áreas, especialmente de Business Intelligence (BI) desenvolveram-se rapidamente e são consideradas como sendo das de uso mais significativo nas tecnologias da informação. O Business Intelligence pode ser definido como sendo uma categoria ampla de aplicações e tecnologias para recolha, armazenamento, análise e disponibilização de acesso a dados para ajudar os utilizadores nas empresas a tomarem as melhores decisões de negócio (Negash et al., 2003).

Considerando que, no passado, o mercado de Business Intelligence foi estritamente dominado por ferramentas comerciais, os tempos recentes foram caracterizados pelo nascimento de soluções open source. Isto representa uma enorme vantagem competitiva, no entanto, a escolha de uma suite open source BI é um grande desafio.

O open source começa hoje a ser visto com outros olhos por parte das empresas, oferecendo cada vez mais garantias e disponibilizando óptimas soluções para o ambiente empresarial, contribuindo para uma grande redução dos custos a nível tecnológico.

O problema principal que este trabalho pretende resolver é levar o Business Intelligence para as empresas, sem custos.

O foco deste trabalho é assim a selecção e implementação de uma suite open source de Business Intelligence, numa empresa de média dimensão, tentando demonstrar que estas soluções são viáveis neste tipo de empresas, pois não existem custos de aquisição, embora requeiram algum esforço e investimento na sua implementação.

1.1. Principais contribuições deste trabalho

Tendo em vista a contribuição para o avanço da ciência e do conhecimento, foi identificado o problema, para cuja solução se pretende contribuir. Para a questão da necessidade de rentabilizar e otimizar a competitividade das PMEs, e credibilização das soluções open source, as principais contribuições deste trabalho são:

- Mostrar que as suites open source BI são uma alternativa a soluções comerciais para a utilização em PMEs.
- Estudo comparativo de diversas suites open source BI.
- Apoiar os processos de implementação e desenvolvimento de suites open source BI em contexto empresarial.
- Avaliar o desempenho de soluções BI open source num ambiente real.
- Mostrar o open source como uma solução credível no contexto empresarial.
- Incutir e divulgar a importância dos conceitos de BI.
- Melhorar e facilitar o processo de decisão numa PME, intervindo sobre áreas específicas.
- Contribuir para o aumento da competitividade numa PME, sem um aumento dos custos.
- Divulgar uma metodologia de comparação de software open source.

São estas, de uma forma geral, as principais contribuições que foram dadas com o desenvolvimento deste trabalho.

De salientar também que foram publicados dois artigos em conferências internacionais com processo de selecção pelos pares:

- “Analysis of Open Source Business Intelligence Suites”, CISTI’2013, António Marinheiro e Jorge Bernardino
- “OpenBRR Evaluation of an Open Source BI Suite”, C3S2E13, António Marinheiro e Jorge Bernardino

1.2. Estrutura do relatório

Este relatório está estruturado em oito capítulos. No actual capítulo, introdutório, procura-se descrever o problema, as principais contribuições deste trabalho e a estrutura do relatório.

O estado da arte, encontra-se descrito ao longo dos capítulos 2 e 3, divididos respectivamente em Open source e Business Intelligence. Assim, o capítulo 2, descreve o software open source, a sua história, o seu impacto nas empresas, sendo dados exemplos de software desta natureza usados diariamente pelo mundo inteiro. No capítulo 3 é feita uma descrição do Business Intelligence, quais os seus componentes, quais os benefícios na sua utilização, bem como das principais dificuldades na sua implementação.

No capítulo 4, é feito um estudo sobre as principais suites open source Business Intelligence existentes no mercado, e quais as características de cada uma. É também descrita a metodologia de avaliação das diversas suites, e realizada a sua comparação.

O capítulo 5 descreve como foi realizada a análise prática das suites, apresenta um método para avaliação de software open source, e demonstra a sua avaliação e consequente escolha da suite a implementar num ambiente empresarial.

No capítulo 6 é apresentada a empresa onde foi realizado e executado o estudo, quais as diversas questões que se pretende que a suite responda, e é apresentado e descrito o modelo dimensional da data warehouse que suporta a extracção de informação, sendo justificadas as opções tomadas.

No Capítulo 7 é descrito como foi realizada a implementação, desde a instalação da suite e softwares acessórios, como foi realizado o ETL (Extract, Transform, Load), e são mostrados exemplos da extracção de informação que responde aos requisitos.

Por fim, no capítulo 8 são apresentadas as principais conclusões deste trabalho, e é descrito algum trabalho futuro.

Existem ainda dois anexos que incluem os dois artigos científicos realizados ao longo deste trabalho, publicados e apresentados em conferências de âmbito internacional.

2. SOFTWARE OPEN SOURCE

Um software open source, também designado por código aberto, baseia-se na partilha de informações, permitindo o uso de tecnologias sem pagamento de licenças, em oposição ao software proprietário.

Um software open source pode ser copiado, executado, distribuído, modificado e aperfeiçoado por todos os seus utilizadores, segundo a definição da OSI (Open Source Initiative) (Open Source, 2014).

Actualmente é usual entender-se que estando perante open source estamos perante software livre. Embora os princípios sejam os mesmos, não são a mesma coisa. Estando perante software livre, há quem entenda ainda que estamos perante software gratuito.

Estes conceitos serão abordados de seguida, pois o software destas diferentes categorias é usado cada vez mais por pessoas, empresas e instituições com vista a uma redução de custos do seu licenciamento. Na figura 2.1. podemos ver o logótipo da OSI, que identifica um software open source.



Figura 2.1. Logótipo da OSI

2.1. A história do software open source

Nas décadas de 50 e 60, o desenvolvimento de software era realizado nos meios académicos, que de forma cooperativa mantinham e distribuían os sistemas que desenvolviam. O código fonte destes projectos era distribuído pela necessidade de correcção de *bugs* ou erros, ou de adequar o seu funcionamento.

Em 1964 os Laboratórios Bell, em conjunto com a American Telephone and Telegraph (AT&T), perante a decisão de usarem um sistema operativo de terceiros ou desenvolverem o seu próprio sistema operativo, optam pela segunda opção e, unindo-se à General Electrics e ao MIT (Massachusetts Institute of Technology), criam o Multics (Multiplexed Information and Computing Service) (Multics, 2014).

Em 1969 a AT&T retira-se do projecto por prever que este demoraria bastante tempo e levaria a custos altíssimos. Os programadores Ken Thompson e Dennis Ritchie juntam-se assim a outros membros e decidem criar uma versão simplificada do Multics, a que em forma de sátira vêm a atribuir o nome de Unics (Uniplexed Information and Computing Service), mais tarde mudado para Unix.

Na década seguinte Thompson e Ritchie publicam um artigo sobre o Unix, incentivando a cópia do seu sistema por parte das universidades. Visto não ser a informática o foco da AT&T, decidiram fornecê-lo sem restrições, incentivando o seu uso, estudo e reestruturação. Iniciam-se assim os encontros para discutir o Unix, compartilhando o código fonte e os estudos para a sua melhoria, o que deu origem a que surgissem desde logo várias versões do sistema.

Em 1984 a AT&T, já após o TCP/IP se ter tornado um padrão da Internet, é dividida em várias companhias, criando uma sucursal para a informática, o que levou a que o Unix deixasse de ser uma plataforma partilhada. Nesta altura já eram comuns os computadores pessoais com sistemas operativos proprietários.

Ainda em 1984 Richard Stallman inicia o projecto GNU (acrónimo recursivo de “GNU is Not Unix”) por achar que a comunidade de programadores que trabalhavam em conjunto tinha sido praticamente extinta com o desaparecimento dos softwares open source. A sua ideia foi então criar um sistema operativo em que qualquer pessoa tenha privilégios e o possa usar, copiar, alterar e distribuir, juntando novamente uma comunidade de programadores. A figura 2.2. ilustra o logótipo adoptado para identificar o GNU.



Figura 2.2. Logótipo do GNU

Para que a migração para o novo sistema fosse facilitada, Stallman sabia que este teria que ser compatível com o Unix, que estava já entretanto protegido por uma licença copyright, que defende os direitos de autor. Assim, ele desenvolve uma licença a que chamou Copyleft, que permite o uso referido a um programa, desde que este mantenha as mesmas opções. O projecto GNU começa então a integrar programas livres, como o compilador gcc (GNU c compiler) e o editor de texto Emacs (Stallman, 1998).

Em 1985 ele cria então a FSF (Free Software Foundation) (FSF, 2014) a fim de desenvolver e manter o sistema GNU, cujo financiamento resultava da venda de CD's de software livre. Foi desenvolvido então o interpretador de comandos BASH (Bourne Again Shell), cujo nome é uma alusão a Stephen Bourne, criador do Shell do Unix. A FSF paralelamente ao desenvolvimento do sistema GNU possibilita ainda aperfeiçoar diversos programas que podiam ser executados em sistemas Unix proprietários. Só no início da década seguinte o sistema GNU é dado como praticamente pronto, faltando desenvolver o seu kernel (núcleo), que possibilitaria a comunicação hardware/software.

Em 1991, Linus Torvalds, então estudante da Universidade de Helsínquia, na Finlândia, desenvolve um kernel compatível com o Unix utilizando o gcc, baseado num sistema operativo desenvolvido para fins educacionais chamado Minix, que tomou o nome de Linux. Linus decidiu publicar na Internet o código do seu kernel para que este fosse estudado e aperfeiçoado por uma comunidade mais vasta. Em 1992 o kernel Linux foi integrado no GNU completando assim este sistema, que se passou a chamar GNU/Linux, actualmente apenas conhecido como Linux.

Ainda nesta década há um marco que define aqueles que são os princípios do open source. Eric Raymond com a publicação do artigo “A Catedral e o Bazar” (Raymond, 1997) influencia a Netscape a disponibilizar o acesso ao código-fonte do seu navegador, hoje conhecido como Firefox. Raymond, a Netscape e outras empresas e programadores começam então a procurar formas para atrair para os princípios do open source empresas que desenvolvem software do tipo comercial.

A publicação de Raymond, que foi apresentada no Linux Kongress (realizado a 27 de Maio de 1997 em Wurzburg, na Alemanha), é considerada como sendo o manifesto do movimento open source, onde inspirado por Linus Torvalds, consagrou a frase “Havendo olhos suficientes, todos os erros são óbvios” (Raymond, 1997).

2.2. Free Software vs Open Source

Embora se possa pensar o contrário, não se refere ao mesmo quando se utilizam os termos Open Source e Free Software, pois há diferenças entre ambos, como iremos ver nesta secção.

Richard Stallman criou em 1985 a Free Software Foundation e delineou os princípios éticos do movimento, que se baseia em quatro liberdades: o software deve ser livre para ser modificado, executado, copiado e distribuído (Free software, 2014)

Um software é livre (Free Software) se os utilizadores tiverem as liberdades descritas. Isto é, um utilizador deve ser livre para distribuir cópias dum software, com ou sem modificações, gratuitamente ou cobrando pela distribuição, para qualquer pessoa, em

qualquer lugar. O ser livre segundo Stallman significa que ninguém tem que pedir permissão ou pagar pela permissão, desde que esteja na posse do software.

Para que a liberdade de realizar modificações e de publicar versões alteradas tenha significado, deve ter-se acesso ao código-fonte do software, sendo esta uma condição necessária para que o software seja livre.

No entanto Stallman foi sempre uma pessoa tida como tendo um discurso muito politizado e de alguma forma radical, nomeadamente na condenação do software proprietário.

Eric Raymond, fazendo parte de uma nova geração de programadores, e não concordando essencialmente com o discurso de Stallman, criticou-o abertamente na publicação já referida “A Catedral e o Bazar” (Raymond, 1997), em que considera o modelo de desenvolvimento GNU como se fossem catedrais, monumentos sólidos, desenvolvidos a partir de um grande planeamento. Já o modelo adoptado por Linus Torvals é comparado a um bazar, com uma dinâmica bastante descentralizada. As alterações do software são rapidamente distribuídas pela comunidade, que analisa e avalia, sendo que as melhorias são rapidamente adoptadas e as falhas são também rapidamente identificadas.

Stallman tenta até hoje deixar claro que o free de “free software” não significa grátis mas livre. Esta confusão tornou-se assim na justificação que faltava para uma rotura com a FSF (Free Software Foundation).

Assim, em 1998, Eric Raymond junto com outras personalidades fundou a Open Source Initiative (OSI), que adoptou o termo open source para se referir ao software livre, com uma postura mais pragmática relativamente à adopção de software de código aberto como uma solução viável, e menos ideológica que a da Free Software Foundation.

As quatro liberdades idealizadas por Stallman são aceites como recomendações práticas. Porém, para que uma licença de software seja considerada de código aberto pela OSI, deve atender aos dez critérios da Definição de Código Aberto (OSI, 2014), que inclui a livre redistribuição, permissão de trabalhos derivados, não discriminação e distribuição da licença entre outros.

As diferenças de princípio entre os dois movimentos não são muito substanciais, sendo-o mais ideologicamente. Os defensores do open source dizem que o termo que adoptaram fez com que os empresários percebessem que o software livre também pode ser comercializado.

Sobre o assunto, na Wikipédia pode mesmo ler-se que “para o movimento do software livre, que é um movimento social, não é ético aprisionar conhecimento científico, que deve estar sempre disponível, para assim permitir a evolução da humanidade” (Software Livre, 2014). Já o movimento pelo open source, defende que o software desse tipo traz

diversas vantagens técnicas e económicas. O open source surgiu para levar as empresas a adoptarem o modelo de desenvolvimento de software livre.

Uma vez que as diferenças são essencialmente ideológicas, as duas correntes unem-se em diversas situações, ou são mencionadas pela sigla FLOSS (Free/Libre Open Source Software).

2.3. Categorias de licenças de software

A maioria das licenças usadas na distribuição de software livre são licenças copyleft. As licenças copyleft são uma versão das copyright (que proíbem que alterações e cópias sejam efectuadas sem autorização do autor), mas definem explicitamente as condições sob quais as cópias, alterações e distribuição podem ser realizadas por forma a garantir as liberdades de modificar e distribuir o software assim licenciado. Podemos ver os logótipos do copyright e copyleft na figura 2.3., sendo o do copyleft o do copyright espelhado.



Figura 2.3. Logótipos do Copyright e do Copyleft

O termo copyleft, como já referido foi criado por Stallman para o projecto GNU, e surge como uma brincadeira com o termo copyright. Flores (Flores, 1988) diz mesmo que “copyleft por sua vez está formado por duas palavras: copy (cópia) e left que pode ser traduzido tanto por esquerda como pelo particípio do verbo leave (deixar, abandonar), de jeito que poderíamos traduzir copyleft por: deixai copiar”.

Há várias categorias de licenças para a distribuição do software livre (Categories of free and nonfree software, 2014), cuja distinção é efectuada pelo grau de liberdade dado ao utilizador. Nada impede que cada interessado crie a sua própria licença desde que cumpra as quatro liberdades básicas, agregando ou não uma cláusula de Copyleft. Destacamos algumas das mais populares.

- GPL – Significa GNU General Public Licence e é a licença que acompanha o software distribuído pelo projecto GNU, incluindo o sistema operativo Linux. Esta licença não só dá a liberdade para a alteração e distribuição do software por

ela protegido, como impede que este seja integrado em software proprietário (GPL, 2014).

- Debian – A licença Debian, chamada de Debian Free Software Guidelines (DFSG) é parte do contrato social celebrado entre a Debian e comunidade de utilizadores open source. Esta licença cumpre os critérios da GPL, mas acrescenta outros onde se salienta a proibição de “contaminar” outro software, ao contrário da GPL que o faz (Debian, 2014).
- Open Source – A licença da Open Source Initiative é derivada da Licença Debian, mas com as menções à Debian removidas (OSI, 2014).
- BSD – Esta licença cobre as distribuições da Berkeley Software Distribution, além de outros. Esta é uma licença considerada permissiva, por colocar poucas restrições quanto ao seu uso, alterações e distribuição. Não há, por exemplo, obrigações quanto à distribuição do código-fonte, e este pode mesmo ser incluído em software proprietário (BSD, 2014).
- X.org – É uma licença do Consórcio X, que distribui o X Windows System, mas que não adere ao copyleft. Há distribuições com esta licença que são software livre, mas outras que não o são (X.org, 2014).

Além deste tipo de licenças, destacam-se as licenças de Software Semi-Livre, em que o software não é livre, mas é concedida permissão ao uso, cópia, distribuição e alteração, desde que seja feito sem o propósito de auferir lucros (Software semi-livre, 2014)

Muito utilizados são também as definições Freeware (Freeware, 2014) (significando software gratuito mas não necessariamente livre como o Free Software) e Shareware (Shareware, 2014) na distribuição de programas, permitindo distribuição, com pagamento da licença no caso de Shareware, mas que geralmente não permitem a distribuição do código-fonte do software, não podendo por isso sofrer alterações.

Na secção seguinte iremos abordar o open source nas empresas, as suas vantagens e desvantagens.

2.4. O open source nas empresas

A implementação de software em qualquer empresa é um processo que deve ser sempre ponderado com uma série de factores que devem ser tidos em consideração. Licenças de software, formação de utilizadores, suporte a dúvidas ou problemas, o custo da sua manutenção, e essencialmente as vantagens produtivas que daí possam advir, são os factores que filtram os tipos de software a adoptar. Normalmente o factor decisivo nesta tomada de decisão é o custo das licenças. No entanto o desconhecimento de alternativas leva muitas vezes as empresas a optar por uma má decisão.

O crescimento do software open source nos últimos anos vem abrir uma nova janela de oportunidade para empresas, instituições, utilizadores e fornecedores de hardware e software.

Embora ainda seja encarado por muitos como uma “moda passageira” que se esvanecerá a mais ou a menos curto prazo (Baptista et al., 2004), deve ser hoje encarado como uma opção credível, criadora de oportunidades de elevado potencial, que deverá ser levada em conta por qualquer decisor ou técnico envolvido de alguma forma com as tecnologias de informação.

A União Europeia tem nos últimos anos sido uma das principais dinamizadoras da adopção de software open source nas suas instituições e também entre os estados membros (Portugal e a realidade europeia, 2012). Em Portugal, esta adopção tem sido lenta, mas as empresas começam a aderir a esta oferta.

Uma aposta cada vez maior das universidades no desenvolvimento de software open source tem levado a uma procura maior por parte das empresas para este tipo de software para diversas soluções como software de escritório, servidores aplicativos, servidores de email, bases de dados, aplicações de gestão de conteúdos, aplicações de gestão de clientes, etc (Oliveira, 2012).

Uma pesquisa realizada em 2010 pela empresa Zenoss (Watters, 2010) e apresentada numa conferência sobre administração de sistemas numa comunidade open source, indicava que 98% das empresas usam de alguma forma algum tipo de software open source. É necessário levar em conta a comunidade em que esta pesquisa foi realizada, mas tal estudo reflecte algo que é inegável. Neste estudo a flexibilidade e os baixos custos foram apontados como o principal factor de escolha de open source, já por outro lado a fraca documentação e a falta de apoio são a sua principal recusa na adopção.

Na União Europeia, empresas como a Airbus, a EDF ou o Deutsche Bank são exemplos de grandes empresas que aderiram ao open source, tal como em Portugal a Zon, a Sonaecom, a Sumol+Compal, a Ordem dos Advogados ou a Egor (Casos de sucesso, 2014).

A nível de instituições estatais tal medida implicaria uma grande redução de custos, e um exemplo a ser seguido é o do Brasil, em que desde há vários anos vários sectores do Estado utilizam software livre, e estabeleceram o seu uso como objectivo estratégico para todo o governo (Kuhn, 2005). Na Europa, também a Rússia assinou um plano para alterar o software das autoridades federais para software livre até ao final de 2014 (Carlos, 2011).

Tais medidas não significam tomar partido pelo software livre ou pelo software proprietário de um ponto de vista fundamentalista, como se um fosse o bom e o outro

fosse o mau. No entanto estes podem e devem coabitar no melhor interesse das pessoas, empresas e instituições envolvidas.

Um grande factor contra o software proprietário é a total obscuridade em termos de garantias de código e a dependência dos fornecedores para actualizações, suporte, e claro, o custo das licenças. Esta é uma das grandes vantagens do open source, a que levou aliás a Rússia a tomar a decisão já referida. Não é apenas a questão económica que está em causa, embora tenha um grande peso, mas é acima de tudo a segurança que a independência relativamente a fornecedores propícia.

Mas nem tudo no open source são vantagens, há pontos que não são tão vantajosos na adopção destas soluções. Como principais vantagens para o open source podem-se referir (Sinfic, 2008):

- O baixo custo ou nulo das licenças.
- As actualizações permanentes.
- A troca de experiências, dúvidas e informações numa comunidade sobre determinado produto.
- A facilidade de dividir uma aplicação em módulos, utilizando só os necessários, bem como realizar a adaptabilidade necessária.
- Integração com outras ferramentas, interligando os componentes disponíveis através da reutilização de código.
- Descontinuidade de determinado software ou de determinadas versões por parte das empresas de software proprietário.
- Independência tecnológica e/ou de um único fornecedor.
- Possibilidade de auditabilidade dos sistemas.
- A menor ameaças de vírus, *trojans* e *worms*, bem como ameaças de segurança.

Como principais desvantagens, destacam-se:

- Produtos pouco amigáveis para a maioria dos utilizadores, e muito dependentes de especialistas para programarem novas funcionalidades.
- Maior dificuldade de implementação, nem sempre com ferramentas gráficas que facilitem esse trabalho.
- Falta de suporte credenciado para uma aplicação, dependendo normalmente da comunidade de utilizadores para resolução de problemas.

- Inexistência de uma versão estável do produto, pois este está sempre em evolução.
- Inexistência de vínculos contratuais.

Na secção seguinte iremos dar alguns exemplos de programas de software open source utilizados em diversos domínios.

2.5. Exemplos de software open source

Há muitos exemplos notáveis de software open source no mercado, com uma grande representatividade, dos quais se destacam alguns, contudo estando todos os dias a surgir novas soluções. A tabela 2.1 mostra uma lista com exemplos em diversos domínios, embora não seja uma lista exaustiva, pois existem muitos programas que não estão referenciados (Lista software open source, 2014).

Tabela 2.1 – Exemplos de software open source em diversos domínios

Domínio	Solução
Modelação de dados	Argo UML, Open Ameos, Jude
Antivírus e manutenção	ClamWin, MemTest
Partilha de ficheiros	Shareaza, Freenet, Azureus, BitTorrent, eMule, LimeWire, Ares, Deluge
Desenvolvimento	Euphoria, Code::Blocks, Dev-C++, Dev Pascal, Eclipse, Free Pascal, Icon, Lazarus, Lua, MinGW, Netbeans, Perl, Python, Quincy, Ruby, Ruby on Rails, Sharp Develop, Unicon, CA-Realia, qt
Editores de Imagem	Paint.NET, Inkscape, Gimp
Modeladores 3D	Blender, Wings 3D
Editores de Som e Imagem	Aimp2, Audacity, Virtualdubmod, MPlayer, Media player classic, VLC, Songbird, Miro, Lmms
Editores de Texto e HTML	AbiWord, LibreOffice, Nvu, LaTeX, Scribus, Notepad++, Vim, Emacs
Emuladores	Stella, O2EM, MEKA, Gens, ZSNES, Finalburn Alpha, Project64, Visual Boy Advance, Nintendo DS
Gestores de Download	Free Download Manager, TrueDownloader, Orbit, Millweed, Tor
Mensagens instantâneas	Pidgin, aMSN, Miranda IM, Mercury IM
Jogos	FreeCol, LGames, SpaceShooter, Frets on Fire, Battle for Wesnoth, UFO Alien Invasion, Quake 1, 2 e 3
Navegador, E-mail e Calendário	Google Chrome, Mozilla Firefox, Mozilla Thunderbird, Mozilla Sunbird, K-Meleon, SeaMonkey
Utilitários	1-4a Rename, 7-zip, Filzip, Vorbis, wxMusik, Filezilla
Office	OpenOffice, LibreOffice, Crimson Editor, DeskLight
Software Científico	SciLab, Octave, Maxima, FreeMat, JMathLib
Gestão de Documentos	GED Maarch

Pode ser encontrada informação permanente sobre este tipo de produtos em diversos repositórios online, de onde se destacam:

- <http://sourceforge.net> – um dos repositórios mais completos sobre aplicações open source, com actualização diária.
- <http://opensource.org> – o site da Open Source Initiative promove os produtos por si certificados.
- <http://www.w3.org/Status> - o site da World Wide Web Consortium disponibiliza também uma grande lista de produtos para diversos domínios.

Neste capítulo fizemos um resumo da história do open source, vimos a diferença entre free software e open source, abordámos os principais tipos de licenças, enquadrámos o open source na realidade das empresas e fornecemos alguns exemplos de software open source em diversos domínios. No capítulo seguinte iremos abordar o Business Intelligence.

3. BUSINESS INTELLIGENCE

Business Intelligence (BI) é um termo proposto e promovido por Howard Dresner da Gartner Group em 1989. Ele descreve-o como um conjunto de conceitos e métodos para melhorar a tomada de decisão empresarial recorrendo a sistemas de apoio baseados em factos (Power, 2007).

Na altura, a indústria de software falava em acrónimos como Sistemas de Apoio à Decisão e Sistemas de Informação Executiva, e Howard Dresner procurava um conceito que elevasse o debate a outro patamar. Este explica que não tinha intenção de colocar o foco do lado tecnológico apenas mas principalmente do lado do negócio.

O Business Intelligence surge assim como um conjunto de conceitos e métodos para auxiliar e aumentar a capacidade de tomada de decisão nas organizações, transformando os seus dados em conhecimento (Negash, 2004).

Já para (Sezões et al., 2006), o BI é um conceito que engloba um vasto conjunto de aplicações de apoio à tomada de decisão que possibilitam um acesso rápido, partilhado e interactivo da informação, bem com a sua análise e manipulação; através destas ferramentas, os utilizadores podem descobrir relações e tendências e transformar grandes quantidades de informação em conhecimento útil.

Em (Sezões et al., 2006) pode ler-se ainda que os sistemas de BI têm em comum o seguinte conjunto de objectivos fundamentais:

- Acesso a dados fiáveis – a fiabilidade dos dados, a sua integração e compreensão entre diversas áreas é fundamental a um consciente exercício de gestão;
- Aumento da transparência e compreensão do negócio – a disponibilização de conhecimento em tempo real (o “quê”, o “quanto”, o “quando”, o “onde” e o “como”) permite aos gestores e decisores ter uma perspectiva das áreas que devem controlar com total transparência e aumentar a sua capacidade de compreensão (do “porquê”);
- Suporte para a tomada de decisão – só uma compreensão oportuna da realidade pode permitir efetuar decisão eficazes; como tal, o conhecimento produzido pelos sistemas de BI, potenciados pelas tecnologias de comunicação actuais, deve suportar e justificar as medidas tomadas pelos vários intervenientes no processo de gestão.

Assim, conforme também já referido no Capítulo 1, as empresas têm actualmente que antecipar mudanças, encontrar soluções e adaptarem-se obtendo vantagens competitivas face à concorrência em momentos adversos.

As empresas devem conhecer os seus pontos fortes e fracos, e também conhecerem os dos seus concorrentes, e terem facilmente acesso e actualização desta informação, traduz-se em vantagem competitiva para o próprio negócio.

Com a exigência e complexidade tecnológica e concorrencial dos dias de hoje, as empresas estão obrigadas assim a prestar uma atenção redobrada ao processo de controlo interno, de desempenho e aos sistemas de suporte à decisão.

Nas secções seguintes, abordaremos a diferença entre dados, informação, conhecimento e decisão. Também serão estudados os componentes do BI, os benefícios na sua utilização e as dificuldades na sua implementação.

3.1. Dados vs Informação vs Conhecimento vs Decisão

Dados são factos isolados, são elementos que isoladamente não têm qualquer utilidade. Até que estes sejam processados e organizados de forma a possibilitar a sua compreensão, não são informação.

Informação, segundo (Galliers, 1987) é um conjunto de dados que, quando fornecido na forma, conteúdo e tempo certos, melhora o conhecimento da pessoa que o recebe, ficando por isso mais habilitada a desenvolver determinada actividade ou a tomar determinada decisão. Isto é, os dados transformam-se em informação quando se lhes acrescenta significado, como a contextualização, categorização, cálculo, correcção ou condensação (Amaral et al., 2007).

Já o conhecimento reside nas pessoas, e só ele permite decisões acertadas. Em (Davenport et al., 1998) é referido que enquanto existirem dados e informações, obtemos o conhecimento em indivíduos ou grupos ou em rotinas das organizações.

Para que a informação se transforme em conhecimento, é necessário trabalho humano como a realização de comparações entre informações, analisar as suas consequências, quais as relações entre a informação obtida e o conhecimento acumulado, e reflectir sobre as novas informações. Só assim o conhecimento auxilia a tomada de decisão.

Já a decisão é o processo pelo qual são escolhidas algumas ou apenas uma entre muitas alternativas para as acções a serem realizadas. Esta escolha deve estar baseada no maior número possível de informações e conhecimento para que a decisão escolhida seja a melhor entre as disponíveis (Antonelli, 2009).

A figura 3.1 ilustra este processo, desde os dados até à decisão.

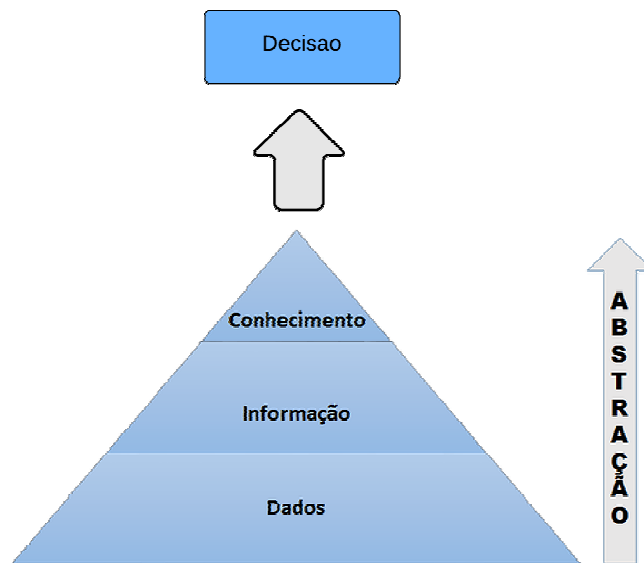


Figura 3.1. Dos dados à decisão (adaptado da fonte:
<http://www.binapratICA.com.br/#!dados-x-informacao/cwk6>)

Na próxima secção abordaremos os diversos componentes do Business Intelligence.

3.2. Componentes do BI

Os componentes do BI consistem em ferramentas que realizam em conjunto o trabalho de armazenamento de dados, análise de informação e prospecção de dados (data mining), como se pode verificar na ilustração da figura 3.2.

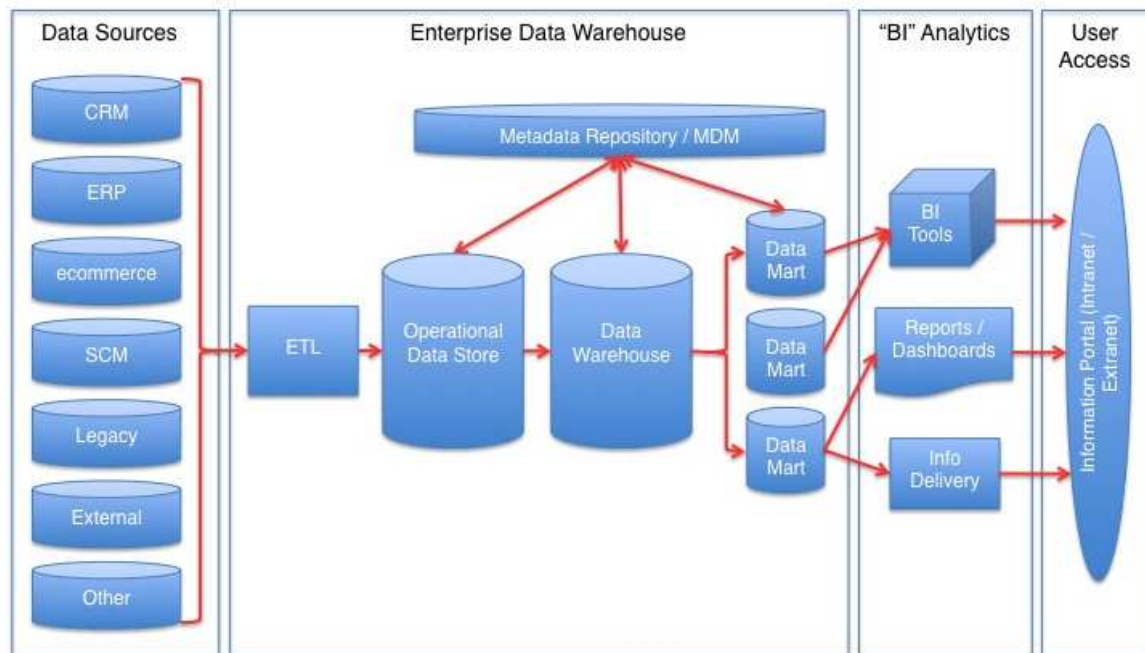


Figura 3.2. Arquitectura típica do BI (fonte: <http://bibuddies.blogspot.pt/>)

Numa arquitectura típica, existem várias origens de dados, como os CRM (Customer Relationship Management) ou os ERP (Enterprise Resource Planning), que servem de base para através do processo de ETL (Extract, Transform, Load) se alimentar uma data warehouse preparada para dar resposta às necessidades de obtenção de informação por parte da organização. Por sua vez, os utilizadores acedem à informação através das ferramentas de BI, reports, dashboards, ou directamente aos dados já tratados que estão armazenados na data warehouse.

Nas secções seguintes iremos abordar os diferentes componentes que constituem o BI, para melhor o compreendermos. Serão então abordados os dados operacionais, o Operational Data Store, ferramentas ETL, Data Warehouse, Data Mining e Visualização dos resultados.

3.2.1. Dados Operacionais

Para que seja possível a implementação de BI numa empresa, é necessário que esta possua dados operacionais armazenados, pelo menos, numa base de dados. Estes dados têm origem no processo operacional da empresa e são guardados pelos sistemas transaccionais usados nas organizações. Estes, são a matéria-prima para o projecto de BI.

Estes dados são gerados no dia-a-dia de trabalho da empresa e podem estar guardados em bases de dados relacionais, ficheiros de texto, folhas de Excel, num ERP, em sistemas desenvolvidos internamente para uma área ou assunto específico, em sistemas legados, etc.

Estas várias origens possuem todas as informações operacionais da empresa desde o registo de clientes, fornecedores, produtos, empregados, facturação, stocks, informação fiscal e financeira, etc. Resumindo, são todas as informações históricas duma empresa armazenadas ao longo do tempo de vida desta, sendo que tal se traduz num grande volume de informação armazenada nas mais diversas origens de dados.

Estando os dados operacionais espalhados por vários repositórios, não é possível efectuar uma análise correcta destes, que normalmente possuem nomenclaturas e termos diferentes em cada fonte, para designar a mesma coisa.

Veremos à frente como se realiza a junção e tratamento destes dados, para que seja possível a sua análise.

3.2.2. ODS – Operational Data Store

O ODS é uma base de dados projectada para integrar dados de diversas fontes, para operações adicionais sobre estes. Os dados do ODS podem regressar para os sistemas operacionais para outras operações, ou para data warehouses para reporting.

Como os dados têm origem em diversas fontes de informação, a sua integração envolve muitas vezes a limpeza, tratamento da redundância e cumprimento das regras de negócio quanto à sua integridade. Um ODS é muitas vezes concebido para conter dados atómicos, com histórico limitado, capturados em tempo real, em oposição a volumes maiores de informação guardados na data warehouse .

Para carregamento de um ODS recorre-se normalmente a virtualização de dados ou extracção, transformação e carregamento (ETL) de dados. Um ODS não é um substituto duma data warehouse, no entanto pode ser uma fonte de dados para esta (Inmon, 1999).

O ODS tem importância no processo de análise de dados pelas ferramentas de BI, pois fornece importantes informações no processo de decisão, devido às suas características de consolidação e integração com as possíveis várias fontes de dados operacionais da organização, de forma centralizada.

3.2.3. Ferramentas ETL

O ETL é usado para designar o processo de carregamento duma data warehouse (no âmbito de BI) tendo como origem várias fontes de dados externas. O processo pode ser usado noutro contexto, como para carregar outras bases de dados que não data warehouses ou outras necessidades.

As ferramentas de ETL permitem efectuar a extracção de dados de diversos sistemas, transformar esses dados segundo as regras de negócio, e efectuar o seu carregamento para data marts ou data warehouses.

O ETL é normalmente considerado uma das etapas mais críticas duma data warehouse, uma vez que envolve a movimentação de dados. A origem dos dados pode ser diversa, desde diferentes bases de dados, a ficheiros de texto, ou outras.

A extracção é a recolha de dados dos sistemas de origem, os dados operacionais, extraíndo-os e passando-os para o ambiente da data warehouse. O ETL pode ser realizado de forma independente dos sistemas operacionais.

A transformação realiza a limpeza, os ajustes e a consolidação, permitindo melhorar a qualidade dos dados e consolidá-los das diversas origens. Nesta fase aplicam-se uma série de regras ou funções aos dados a serem carregados. Sobre algumas origens pode não ser necessária grande manipulação de dados, já noutras os dados necessitam ser trabalhados para um correcto carregamento. A transformação pode passar por exemplo por seleccionar apenas algumas colunas a carregar, ou traduzir dados para uma linguagem uniforme da data warehouse (por exemplo, relativamente ao campo sexo, numa origem feminino pode ser 0, noutra 1, ou F, e na DW ter sido adoptado Feminino).

O carregamento consiste em passar os dados já transformados para o modelo dimensional da DW. Este passo varia consoante as necessidades das organizações, ou das opções de projecto seguidas. Podem ter intervalos de carregamento definidos, actualizar dados acumulados ou acrescentar dados nas dimensões.

Todo este processo ocorre numa área da data warehouse designada por DSA (Data Staging Area), que é uma área de armazenamento, fora dos limites dos utilizadores finais, e que não fornece serviços de pesquisa nem de apresentação de dados.

A gestão de um sistema ETL, exige algumas tarefas específicas de carregamento de dados, planos de backup, verificação da segurança, etc.

Antes de se iniciar um projecto de ETL, é necessário levar em conta alguns requisitos, como (Ribeiro, 2011):

- Requisito de negócio – está clarificado e documentado quais são as necessidades?
- Viabilidade dos dados – foi realizada alguma análise da viabilidade dos dados?
- Latência dos dados – qual o tempo máximo permitido para disponibilização dos dados no sistema de BI?
- Políticas de segurança – quais são as políticas de acesso e segurança dos dados adoptadas na empresa?

O ETL é assim um dos processos do ciclo de vida da data warehouse, sendo considerado um dos mais críticos e morosos, onde é normalmente despendido entre 55% e 80% do esforço de todo o processo de implementação (Inmon, 1997; Demarest, 1997).

3.2.4. Data warehouse e o modelo em estrela

Em (Kimball et al., 2013) é definida uma data warehouse como sendo uma “cópia” dos dados transaccionais especificamente estruturada para consultas e análise.

Segundo (Inmon, 2005), uma data warehouse é um conjunto de dados baseados em assuntos, integrado, não volátil, e variável em relação ao tempo, para apoio a decisões de gestão.

No fundo, através de transacções e processos, uma data warehouse realiza “cópia” de dados registados nas diversas origens de dados da organização, reorganizando essas informações e preparando-as para a realização de consultas, relatórios e análises.

Isto é, trata-se dum armazém de dados com origem nos dados operacionais que suprirá a necessidade das regras para criação de visões estratégicas e de gestão para a administração.

As informações armazenadas numa DW são organizadas pelas áreas importantes para o negócio da empresa, podendo existir vários níveis de detalhe. Para o seu correcto funcionamento, e uma vez que os dados podem ter diversas origens, deve existir uma codificação única para os dados. Os dados contidos numa DW referem-se a um momento específico da informação no tempo. Os dados podem assim ser comparados ao longo do tempo, mantendo um histórico da trajectória da informação durante a sua existência. Não é volátil porque uma DW não sofre alteração nos dados, a não ser a adição de novas informações.

Segundo (Kimball et al., 2013), as principais características para garantir a qualidade dos dados numa DW são a unicidade, evitando a duplicação de informação; a precisão, pois os dados não podem perder as suas características originais; a completude, não gerando dados parciais de todo um conjunto relevante às análises; e a consistência, pois os factos devem apresentar consistência com as dimensões que compõe.

O modelo em estrela é o modelo dimensional típico na implementação numa data warehouse, no qual existe uma tabela central, tabela de factos, à qual estão ligadas as tabelas de dimensão, como ilustrado na figura 3.3.



Figura 3.3. Modelo em estrela (fonte: <http://plotting-success.softwareadvice.com/beginners-guide-to-bi-software-1113011/>)

Na figura 3.3 é ilustrado o esquema em estrela típico de um negócio de vendas a retalho. Mostra os factos das vendas de um produto num determinado dia em determinada loja para um determinado cliente. O refinamento da informação neste modelo é definido pela granularidade da dimensão “Time”.

Neste modelo, os dados são mais fáceis de compreender e de navegar, porque:

- O próprio modelo é simples de entender visualmente, com uma tabela central (os factos) e as dimensões à sua volta.
- As pesquisas têm melhor desempenho, pois os dados nas dimensões estão fortemente desnormalizados por forma a termos o menor número de junções entre tabelas;
- Facilmente se podem adicionar novas dimensões ou alterar uma tabela de dimensão.

A tabela de factos é a tabela principal (central) do modelo de negócios, onde as medições numéricas do negócio são guardadas. É constituída por atributos numéricos (factos), e por chaves forasteiras que ligam a tabelas de dimensão. Como é uma tabela que contém uma enorme quantidade de dados (estima-se que mais de 95% do espaço ocupado pela DW), é uma tabela que está fortemente normalizada.

Todas as medições na tabela de factos têm que ter a mesma granularidade (dia, hora, etc), que determina a “precisão” com que poderão ser realizadas as análises.

Neste modelo, há tantas dimensões quantas as vertentes sob as quais se pretende analisar os factos. Como as tabelas de dimensão têm poucos registos (comparativamente com a tabela de factos), são tabelas com muitos atributos (fortemente desnormalizadas), para evitar mais junções mas que contêm os descritores do negócio de forma a permitirem realizar os agrupamentos de pesquisas necessários. Isto é, como já referido, as pesquisas são realizadas mais rapidamente pois existem menos junções (join) que são operações muito dispendiosas do ponto de vista de processamento.

Para elaborar o modelo dimensional, devem ser seguidos quatro passos:

1. Identificar o processo de negócio e modelizar – compreender os requisitos do negócio conjuntamente com os dados disponíveis.
2. Definir a granularidade do processo de negócio – especificar o nível de detalhe que a tabela de factos representa.
3. Escolher as dimensões para cada tabela de factos – representam todas as descrições possíveis que tomam os valores únicos no contexto de cada medição.
4. Identificar os factos numéricos de cada tabela de factos – respondem ao que estamos a medir, dentro da granularidade definida.

Nesta secção abordámos o modelo em estrela numa data warehouse. Na secção seguinte abordaremos outro componente do BI, o data mining.

3.2.5. Data mining

O uso de DW só se torna interessante com a utilização de boas ferramentas de exploração. No processo de BI o data mining (prospecção de dados) é assim um factor decisivo. O data mining é o processo de explorar grandes quantidades de dados à procura de padrões consistentes, como regras de associação ou sequências temporais, para detectar relacionamentos sistemáticos entre variáveis, detectando assim novos subconjuntos de dados (Fayyad et al., 1996).

O modelo da extração do conhecimento pode ser visto na figura 3.4., sendo que o data mining é constituído por quatro fases principais, sendo elas:

1. Fase de preparação – são as actividades que vão desde a criação duma base de dados separada para os dados sujeitos a análise, até ao seu processo de carregamento.
2. Fase de prospecção – é a fase de criação dos modelos de data mining, definição de amostras e selecção de dados.
3. Fase de análise – fase onde é realizada a análise das informações, criando os atributos chave para a análise do negócio. Alguns modelos que podem ser usados nesta fase são a agregação, classificação, padrões sequenciais, e regras de associação.

4. Fase de aplicação – fase em que se utilizam os algoritmos ajustados em situações reais do sistema.

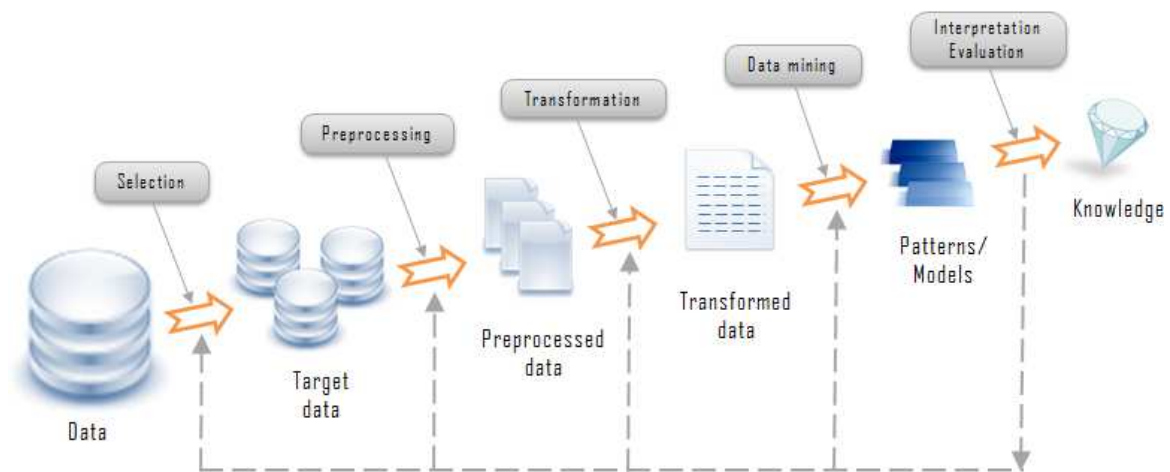


Figura 3.4. Processo da extração do conhecimento (fonte: <http://www.zentut.com/data-mining/what-is-data-mining/>)

O data mining pode oferecer uma alternativa para as empresas descobrirem novas oportunidades de negócio, e traçarem novas estratégias para o futuro, sendo muito relevante nos momentos em que vivemos.

O data mining é então o processo de descobrir padrões e relacionamentos entre dados, partindo de algoritmos estatísticos e matemáticos. Os algoritmos de “associação”, por exemplo, dão-nos respostas com a probabilidade em percentagem e também com a margem de erro. Exemplificando, se analisarmos as vendas realizadas num determinado período, levando em conta o perfil do cliente, podemos obter como resultado que clientes de média dimensão, do segmento de distribuição, localizados na região Sul, têm 77% de probabilidade de comprar determinado tipo de produto, com uma margem de erro de previsão de 12%.

Outros algoritmos, como os de “agrupamento”, permitem classificar os clientes por segmentos, cada um destes com características e comportamentos semelhantes, facilitando a definição de estratégias e acções de marketing.

3.2.6. Visualização dos resultados

A visualização das informações obtidas a partir de todo o processo descrito anteriormente, de forma amigável, ajuda os utilizadores do BI a tomar decisões. Há diversas ferramentas que descreveremos mais à frente que auxiliam a tomada de decisões, como por exemplo ferramentas de reporting e de análise multidimensional dos dados.

Na figura 3.5, podemos observar exemplos de gráficos disponibilizados por essas ferramentas que facilitam a interpretação da informação. É bastante comum, por ser visualmente intuitivo, o uso de gráficos de barras ou de tartes, e o uso de *speedometers gauges*, *bubbles* ou *sliders* para valores de indicadores.

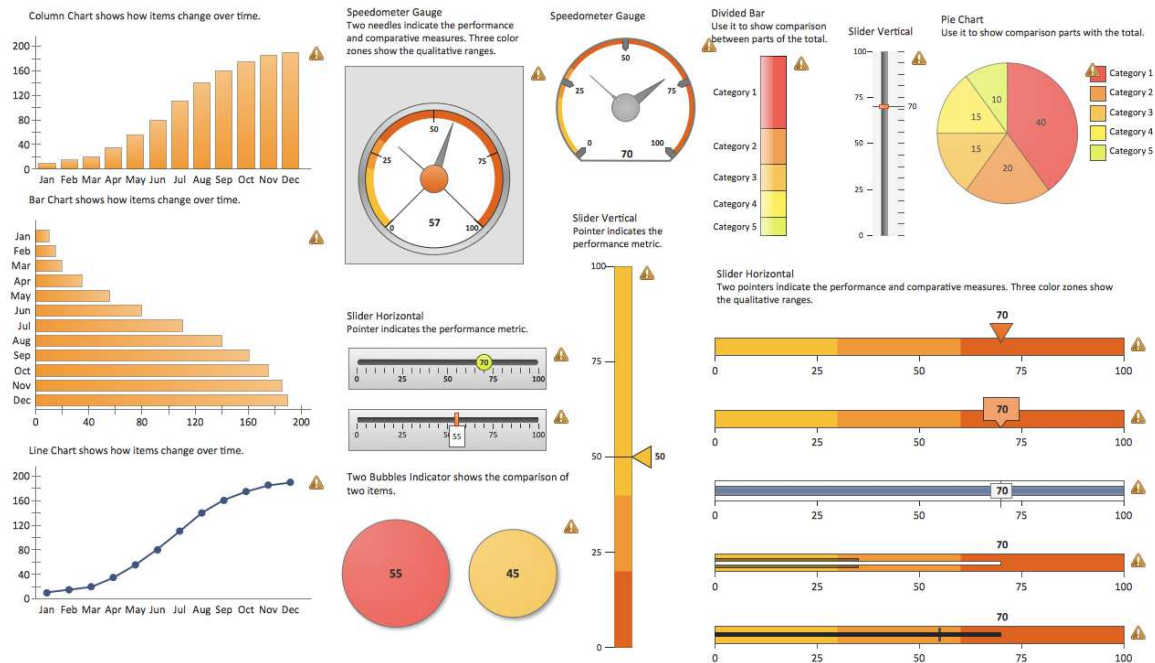


Figura 3.5. Exemplos de gráficos de extração de informação (fonte: <http://www.conceptdraw.com/examples/business-dashboard>)

Nesta secção abordámos os diferentes componentes que constituem o BI. Na secção seguinte abordaremos os benefícios na sua utilização.

3.3. Benefícios na utilização de BI

Os benefícios de utilização de BI para as empresas são enormes, no entanto é importante referir que a análise e levantamento de requisitos realizados, a qualidade e profundidade destes, e a forma como a implementação é efectuada ditam o sucesso ou insucesso do projecto.

Segundo (Linden, 2011), há dez regras clássicas a seguir para implementar com sucesso uma solução de BI numa organização, além de aprender com os erros de outros projectos:

1. Definir os requisitos funcionais;
2. Definir os grupos de utilizadores;

3. Envolver os utilizadores numa fase inicial;
4. Ter apoio da gestão de topo;
5. Identificar os indicadores de desempenho (KPI) requeridos;
6. Garantir a integração e qualidade dos dados;
7. Descobrir que ferramentas de BI já estão disponíveis na empresa;
8. Escolher o software de BI correcto;
9. Limitar o tempo de execução do projecto;
10. Um projecto de BI é um processo em constante evolução, sendo necessária a sua manutenção, actualização, e garantia de qualidade e funcionamento ao longo do tempo.

Quanto aos benefícios, tomando por base um projecto implementado com sucesso, descrito em (Primak, 2008) são referidos:

- Redução de custos com software;
- Redução de custos com administração e suporte;
- Redução de custos na avaliação de projectos;
- Redução de custos com formação dos colaboradores;
- ROI (retorno do investimento) mais rápido para projectos implementados na empresa;
- Maior controlo e menos dados incorrectos;
- Maior segurança da informação;
- Alinhamento da informação estratégica e operacional;
- Facilidade de controlo de acesso e definição de níveis de gestão;
- Maior alinhamento dos utilizadores da empresa;
- Rapidez na informação para tomada de decisões estratégicas;
- Informação consistente em vários locais dispersos;
- Vantagem competitiva.

Estas são as principais vantagens na utilização do BI. Na secção seguinte serão descritas as principais dificuldades.

3.4. Dificuldades na implementação de BI

Na implementação de BI há várias barreiras a serem transpostas por forma a garantir o sucesso do projecto. Muitos erros já foram cometidos, com os quais se deve aprender para evitar insucessos futuros.

Em (Primak, 2008) são referidas algumas das principais dificuldades:

- Dados operacionais estão dispersos e muitas vezes incoerentes com a organização;
- Deficiência dos sistemas operacionais utilizados pelas organizações, que não armazenam dados úteis para futura tomada de decisão;
- A organização não reconhece as necessidades de informações, e só reconhece quando muitas vezes é tarde demais;
- A falta de conhecimento informático dos gestores pode deixar um projecto de BI sem utilidade prática;
- Necessidade de uma boa inter-relação entre a área de negócio e a equipa de implementação;
- As ferramentas técnicas operacionais da actualidade são dispersas e ineficientes, e necessitam de uma reconstrução para serem utilizadas para o BI;
- A obtenção de informações de diversas fontes externas é feita de uma maneira que a relação custo benefício pode não ser favorável;
- Alguns projectos falham como consequência de uma adopção errada de hardware e software;
- Muitos problemas podem ocorrer devido à falta de experiência e conhecimento do fornecedor da ferramenta de BI;
- O tratamento dos dados (ETL) e o armazenamento na DW é um processo que deve ser bem planeado, por ser trabalhoso e complexo, necessitando de profissionais qualificados para garantir o sucesso desta etapa;
- Simples erros na elaboração e desenvolvimento de uma DW podem ser fatais e trazerem resultados negativos ao projecto;
- Dificuldade de realizar o nivelamento entre o BI e a gestão do conhecimento da organização;

- O custo de implementação de um projecto de BI é muito dispendioso.

Kevin Quinn, vice-presidente de Produtos de Business Intelligence e Serviços de Apoio Comercial da Information Builders, refere mesmo que após uma análise realizada a resultados menos positivos de implementação de soluções de BI ao longo dos anos, identificou aquelas que são as quatro piores práticas de implementação do BI (Quinn, 2011):

1. O primeiro erro acontece na escolha da solução de BI, pois os decisores não incluem normalmente os utilizadores da empresa neste processo, e acabam por escolher soluções demasiado complexas para estes. Refere mesmo que apenas 10% dos utilizadores duma organização têm conhecimentos para usar uma ferramenta de BI. Para contornar esta solução devem escolher ferramentas fáceis de usar por todos, especialmente pelos utilizadores não técnicos.
2. O segundo erro identificado, tem a ver com permitir que o Excel se transforme na ferramenta de BI por defeito. Independentemente da qualidade e facilidade da ferramenta, o seu ponto fraco reside na qualidade e coerência da informação gerada, pois os seus processos manuais são muitas vezes origem de erros. A solução pode passar por converter o Excel num front-end do BI, trabalhando sobre dados exactos, pré-formatados e pré-calculados.
3. Outra prática comum é considerar que uma data warehouse resolve todas as necessidades de acesso e distribuição dos dados duma empresa. O problema não é a DW em si, pois esta é fundamental no BI, mas quando esta é considerada a solução para todos os problemas de informação, ou quando se espera que a disponibilidade da DW conduza os utilizadores à informação. As DW não devem ser implementadas sem se conhecer com clareza a necessidade do negócio.
4. Por fim, outro erro comum é a aquisição de software de BI para análises gerais. Os custos mais elevados e a rentabilidade mais reduzida do BI derivam da aquisição duma solução genérica, sem um objectivo específico, o que raras vezes tem impacto positivo no negócio.

Neste capítulo abordámos o Business Intelligence quanto à sua necessidade por parte das empresas, quais os seus componentes, os seus benefícios e a dificuldade da sua implementação.

No capítulo seguinte abordaremos as suites open source de Business Intelligence.

4. SUITES OPEN SOURCE BI

Suite é o termo usualmente utilizado para descrever ferramentas que agregam um conjunto de funcionalidades. Nesta secção serão analisadas as suites que agregam as várias funcionalidades do BI, e que cumprem as necessidades para a implementação numa empresa. Apresentaremos a análise relativamente a cinco suites BI: JasperSoft (Jaspersoft, 2013), Palo (Palo, 2013), Pentaho (Pentaho, 2013), SpagoBI (SpagoBI, 2013) e Vanilla (Vanilla, 2013). Escolhemos estas ferramentas por serem as que actualmente existem com o conceito de suite open source, activas, no mercado.

4.1. Suites estudadas

Nesta secção descreveremos cada uma das cinco suites open source, quanto às suas características e funcionalidades.

4.1.1. JasperSoft

A Jaspersoft afirma-se como fornecendo uma modalidade de self-service à medida das necessidades individuais das empresas (Jaspersoft, 2014). A sua suite Jaspersoft foi criada em 2006, depois de vários anos a empresa ter criado as diversas ferramentas individualmente. A Jaspersoft afirma que fornece o conjunto de BI mais flexível, económico e amplamente implementado no mundo. No seu site, pode ler-se que já foi realizado o download de mais de 14 milhões de cópias do software de código aberto em todo o mundo, 175.000 implementações de produção e que têm mais de 14.000 clientes em 100 países. Afirmando ainda que a suite é regularmente actualizada por uma comunidade de desenvolvimento de mais de 225.000 membros registados. A figura 4.1 representa o logótipo da suite Jaspersoft.



Figura 4.1. Logótipo da Jaspersoft

Possui a distribuição Community, open source, gratuita, e a distribuição Comercial distribuída por três edições (Express, Professional e Enterprise). A distribuição Community é bastante limitada relativamente às distribuições comerciais e é distribuída segundo uma licença GNU GPL. É composta pelos módulos JasperReports Server, JasperReports Library, JasperSoft ETL, JasperSoft Studio e iReport Designer.

O JasperReports Server é um servidor de relatórios “stand-alone” que pode ser incorporado em qualquer aplicação Java. Fornece relatórios e análises que podem ser incorporadas em aplicações Web ou móveis. Fornece informações em tempo real ou agendadas, para a Web, dispositivo móvel, impressora ou correio electrónico numa variedade de formatos. Está optimizado para partilhar, proteger e gerir de forma centralizada os relatórios e análises. Entre as funcionalidades, destacam-se: formatação e visualização de relatórios de forma interactiva, repositório centralizado e seguro, agendamento de geração e distribuição de relatórios e interface personalizável.

Segundo a JasperSoft, o JasperReports Library é o mecanismo mais popular de relatórios open source. É inteiramente programado em Java e é capaz de usar dados provenientes de qualquer tipo de origem e produzir documentos que podem ser visualizados, impressos ou exportados numa variedade de formatos, incluindo HTML, PDF, Excel, OpenOffice e Word.

A integração de dados (ETL – Extract, Transform, Load) é suportada pelo JasperSoft ETL. Permite extrair dados de várias fontes, transformar os dados com base em regras de negócio definidas e carrega-os numa data warehouse ou data mart para análise e extracção de relatórios. Entre as funcionalidades, destacam-se o ambiente de trabalho gráfico, mais de 500 ligações a componentes e o controlo de versões de *jobs*.

O JasperSoft Studio é um ambiente de desenho de relatórios baseado em Eclipse (Eclipse, 2014) para JasperReports e JasperReports Server. Permite criar relatórios a partir de qualquer origem de dados, formatado para visualização no ecrã ou formato de impressão, com exportação para uma ampla gama de formatos. Entre as funcionalidades, destacam-se o ambiente de trabalho gráfico; os modelos de relatórios suportados por temas; integração com JasperReports Server; layouts sofisticados com gráficos, imagens, sub-relatórios e tabelas cruzadas; aceder aos dados através de JDBC, TableModels, JavaBeans, XML, Hibernate, CSV e origens personalizadas; e publicar os relatórios em formato PDF, RTF, XML, XLS, CSV, HTML, XHTML, ficheiros de texto, DOCX ou OpenOffice.

Baseado em NetBeans, o iReport Designer, é uma ferramenta à imagem do JasperSoft Studio, basicamente com as mesmas funcionalidades. Actualmente, a versão mais recente do JasperSoft Studio é a 5.5.1, sendo a última actualização de Fevereiro de 2014. A figura 4.2 mostra um aspecto do iReport Designer, integrante da suite Jaspersoft.

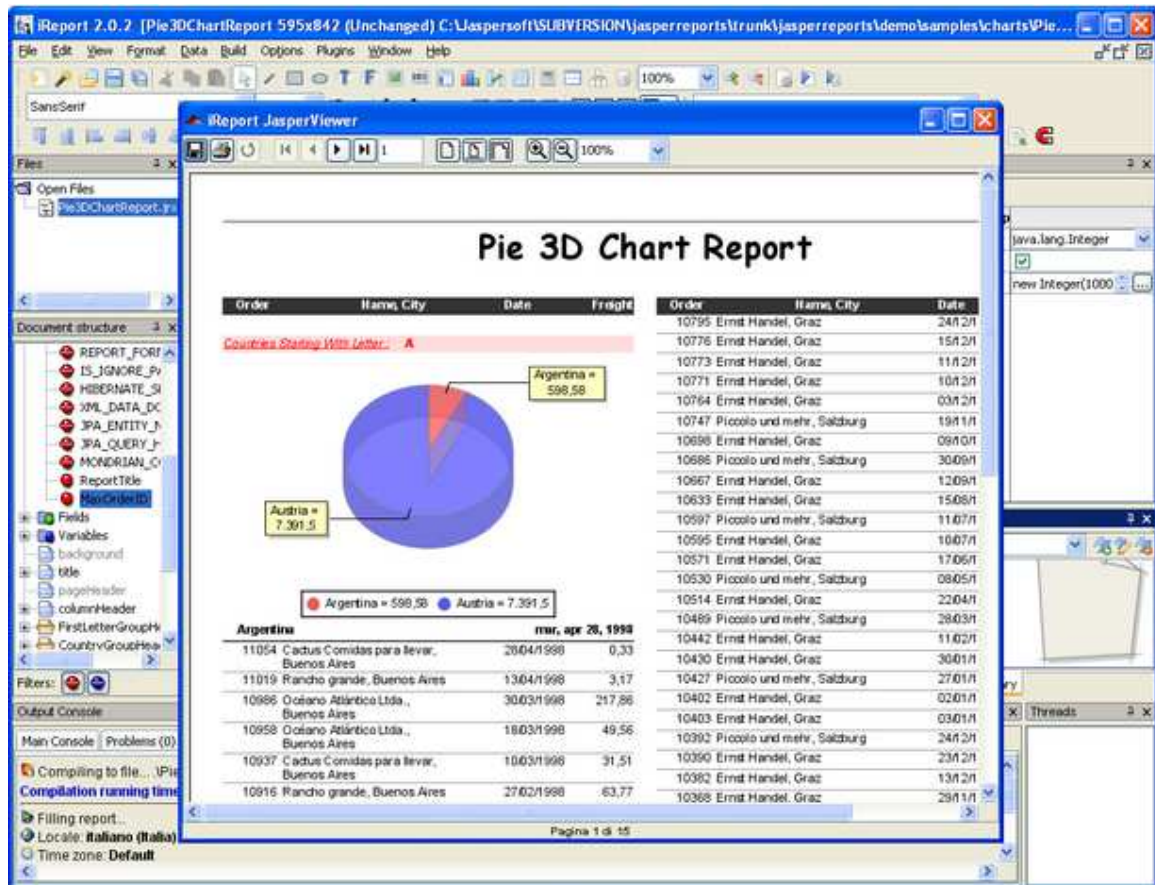


Figura 4.2. Um aspecto do iReport Designer, da suite Jaspersoft (fonte: <http://community.jaspersoft.com/>)

4.1.2. Palo

Palo é como é designada a suite open source desenvolvida desde 2007 pela empresa alemã Jedox AG (Palo, 2014). Esta empresa desenvolve ainda uma versão comercial, com o nome Jedox, que tem normalmente um avanço no código desenvolvido entre 12 a 15 meses, relativamente à Palo. A plataforma é completamente baseada em código open source, totalmente gratuita e distribuída sob uma licença GNU GPL (v2), sendo o seu logótipo o que se encontra ilustrado na figura 4.3.



Figura 4.3. Logótipo da Palo

As características do Palo, agrupam-se em três grupos: Excel Integration (Excel Add-in; OLAP); Web-Client (Web Spreadsheet; Web Pivot; File Manager; User Manager; ETL

Manager); ETL (ETL Server). É ainda composta pelos produtos Palo OLAP, Palo for Excel, Palo Web e Palo ETL.

O Palo OLAP é um servidor multidimensional OLAP, que trabalha em memória, estando os dados organizados em cubos, dimensões e elementos. Esta estrutura reflecte a forma como usualmente os utilizadores pensam o seu negócio. O facto de trabalhar em memória traduz-se numa vantagem relativamente a outros sistemas OLAP com suporte a uma base de dados relacional, que segundo os responsáveis chega a ser 100 vezes mais rápido, permitindo disponibilizar dados em tempo real.

O Palo for Excel é um *add-in* para o Excel e Open Office, que agregado ao OLAP Server Database constitui uma solução completa de BI. Permite integrar no Excel as dimensões relevantes da organização disponibilizadas pelo Palo, permitindo a sua visualização e análise simultaneamente.

O Palo Web combina todos os componentes do Palo, que permite criar análises, planeamento ou relatórios, numa plataforma uniforme Web. Integra o Palo Spreadsheet, o Palo Pivot (realização de pivot tables ad-hoc), o Palo User Manager para gestão de acessos a utilizadores ou grupos, o Palo ETL Manager que cria e controla os projectos de integração de dados e, o Palo File Manager que disponibiliza a cada utilizador ou aplicação espaço para armazenamento de ficheiros, facilitando a integração e colaboração.

O Palo ETL permite realizar a integração, através de ETL, de grandes quantidades de dados a partir de origens heterogéneas. Baseado em ambiente Web, permite facilmente a conversão de um modelo relacional para um modelo OLAP, e realiza o armazenamento em memória necessário para o Drill-through (ligação entre origens de dados), que exhibe detalhadamente através de navegação. A versão mais recente do Palo é a 3.2, de Julho de 2011, mostrando a figura 4.4. um aspecto das opções da suite.

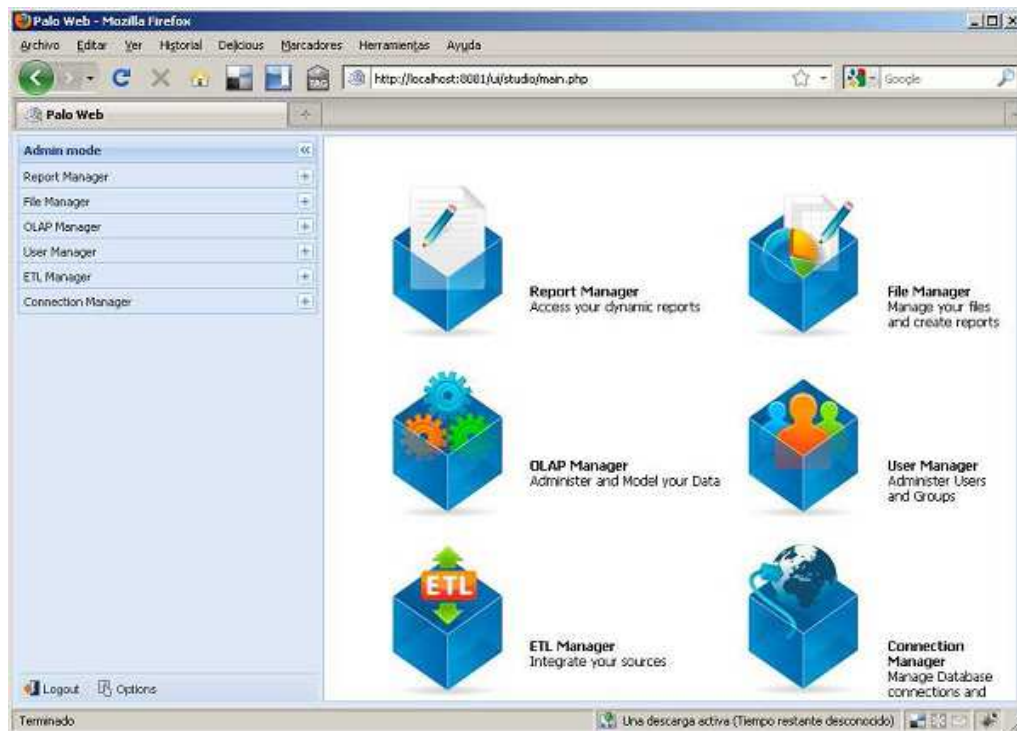


Figura 4.4. Um aspecto da suite Palo (fonte: <http://www.dataprix.com/empresa/productos/palo-suite-jedox>)

4.1.3. Pentaho

A Pentaho foi fundada em 2004 por um grupo de executivos com grande experiência em BI oriundos de empresas como a Business Objects, Cognos, Hiperion, JBoss, Oracle, Rad Hat e SAS (Pentaho, 2014). A suite Pentaho é actualmente das mais usadas e das que maior reputação tem no mercado (Schlegel et al., 2013). A figura 4.5. ilustra o logótipo da Pentaho.



Figura 4.5. Logótipo da Pentaho

Esta suite é composta pelas versões Enterprise e Community. A versão Enterprise é uma distribuição paga que contém todos os recursos, com interfaces amigáveis e de fácil compreensão para os utilizadores finais. A versão Community é open source, gratuita, e fornece praticamente os mesmos módulos, sendo que alguns não têm a interface tão amigável, e têm maior dificuldade de compreensão e implementação. O seu desenvolvimento é feito em Java, e pode ser executada a partir duma Java Virtual Machine. A versão Community, é composta pelos módulos Pentaho Data Integration,

Pentaho Reporting, Pentaho Dashboards, Pentaho Analysis Services e Pentaho Data Mining.

O Pentaho Data Integration, também conhecido como Kettle, permite realizar o ETL dos dados de forma fácil e intuitiva. Suporta uma grande biblioteca de objectos de mapeamento com suporte a diversas fontes de dados, e permite o armazenamento de dados para dimensões de data warehouses, outras bases de dados ou ficheiros. A transformação permite a limpeza de dados através de regras e a migração de dados entre aplicações.

O Pentaho Reporting é composto por duas ferramentas, uma de geração de relatórios, também conhecida como JFreeReport, e outra para geração de metadados, que permite a criação de relatórios ad-hoc via Web. Suporta várias fontes de dados, incluindo relacionais, OLAP e XML, permite desenhar relatórios com suporte por wizard, multi-língua, filtros de visualização interactiva, e exportação para os formatos mais comuns.

Os indicadores de desempenho (KPIs) são garantidos pelo Pentaho Dashboards, que permite a criação de painéis de controlo, e reunir na mesma janela os principais indicadores de um departamento ou de toda a empresa. As métricas são disponibilizadas de forma intuitiva, e permite integração com o Pentaho Reporting e Analysis. Disponibiliza ainda uma monitorização contínua com alertas de notificação de problemas.

O Pentaho Analysis Services, também conhecido como Mondrian, é um motor OLAP, baseado numa arquitectura ROLAP, que pode ser usado com os principais Sistema de gestão de base de dados (SGBD), e possui funcionalidades como camada de metadados, linguagem (MDX, 2014), cache em memória, tabelas agregadas, etc. Suportada num ambiente totalmente Web, permite a criação de relatórios com suporte drag&drop, visualização em gráficos, exportação da informação e visualização multi-dimensional com selecção de métricas e atributos a analisar.

Também conhecido como Weka, o Pentaho Data Mining é um conjunto de ferramentas para prospecção de dados, cujo conjunto de regras de classificação, regressão, associação e algoritmos de clustering permite ajudar a entender melhor o negócio e melhorar o desempenho futuro.

A versão Community é distribuída sob as licenças GNU General Public License versão 2.0 (GPLv2), GNU Lesser General Public License versão 2.0 (LGPLv2), e Mozilla Public License 1.1 (MPL 1.1) [12]. A versão mais recente desta suite é a 5.0.1, disponibilizada em Novembro de 2013. A figura 4.6. mostra um aspecto da suite.

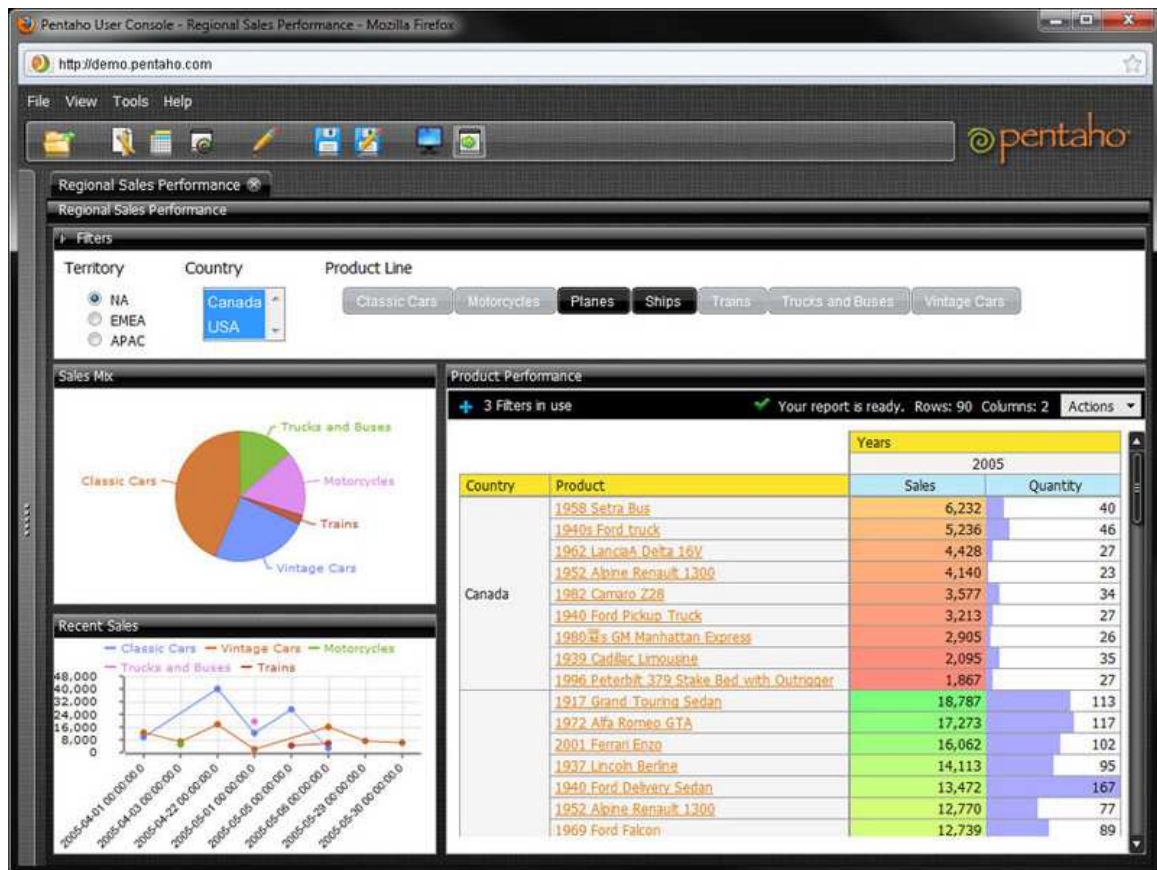


Figura 4.6. Um aspecto da suite Pentaho (fonte: <http://pentaho-bi-suite.blogspot.pt/2013/07/pentaho-community-dashboardcde-basics.html>)

4.1.4. SpagoBI

O SpagoBI é um software para desenvolvimento de projectos de BI em ambientes integrados (SpagoBI, 2014). É suportado pela empresa Engineering (Engineering, 2014) que, de acordo com o seu site, não tem interesse na sua exploração comercial, pois é uma empresa de consultadoria. Para desenvolver a ferramenta a empresa criou a SpagoWorld Initiative em 2006, que mantém a ferramenta com uma comunidade open source, actualmente sob uma licença GNU Lesser General Public Licence (LGPL). Só há uma versão do SpagoBI, gratuita e completa, ao contrário de outras suites que têm versões “community”, com funcionalidades reduzidas, e versões “enterprise”, com as funcionalidades necessárias a uma total integração numa organização. Toda a escalabilidade necessária numa organização é suportada pelo SpagoBI, a nível de arquitectura, funcionalidades e segurança. O seu logótipo está mostrado na figura 4.7.



Figura 4.7. Logótipo da SpagoBI

A ferramenta permite uma visão de integração e não de produto, isto é, possibilita que, por exemplo, a funcionalidade OLAP seja usado pelo Mondrian ou o JPalo, ou que para gerar relatório se use o BIRT (Business Intelligence and Reporting Tools), Jasper ou Business Objects. Relativamente a segurança, o SpagoBI autoriza a criação de regras por grupos de utilizadores para a visualização de dados, e tem total integração LDAP e Active Directory.

Entre as principais funcionalidades, podem-se destacar a geração de relatórios sobre quatro motores distintos (JasperReport, BIRT, Accessible Report, Business Objects), análise multidimensional (OLAP) suportada por três motores (JPivot/Mondrian, JPalo/Mondrian, JPXMLA), ou a geração de gráficos (JFreeChart, HChart, ExtChart) dos mais variados tipos (histogramas, barras, áreas, etc.). Quanto a indicadores de desempenho (KPI) permite a criação, gestão e exploração sobre modelos hierarquizados, através de diferentes métodos, regras de cálculo, limites e regras de alarme. Permite ainda relatórios ad-hoc (que possibilita que os utilizadores criem os seus próprios relatórios, com suporte a multi-página, tabelas simples e cruzadas, gráficos e diferentes tipos de layout), georeferenciação por GEO ou GIS (estabelecendo relação entre dados geográficos e dados da data warehouse), ou pesquisas ad-hoc suportadas por interface Web são outras das funcionalidades disponibilizadas e que constituem uma mais-valia desta suite.

A SpagoBI é uma suite bastante completa que suporta ainda data mining (permitindo descobrir padrões de informação), ETL (Extract, Transform, Load) para carregar e gerir a data warehouse, Mobile BI (disponibilizando componentes específicos para interação com equipamentos móveis, preparados para touch-screen). Destacam-se ainda a facilidade de trabalho colaborativo, integração com Open Office ou MS Office, e consolas de monitorização em tempo real, que podem ser integradas noutras aplicações ou processos.

Este conjunto de funcionalidades torna-a uma suite bastante completa e robusta, totalmente open source, cuja versão mais recente é a 4.1, disponibilizada em Dezembro de 2013. A figura 4.8. mostra um aspecto desta suite.

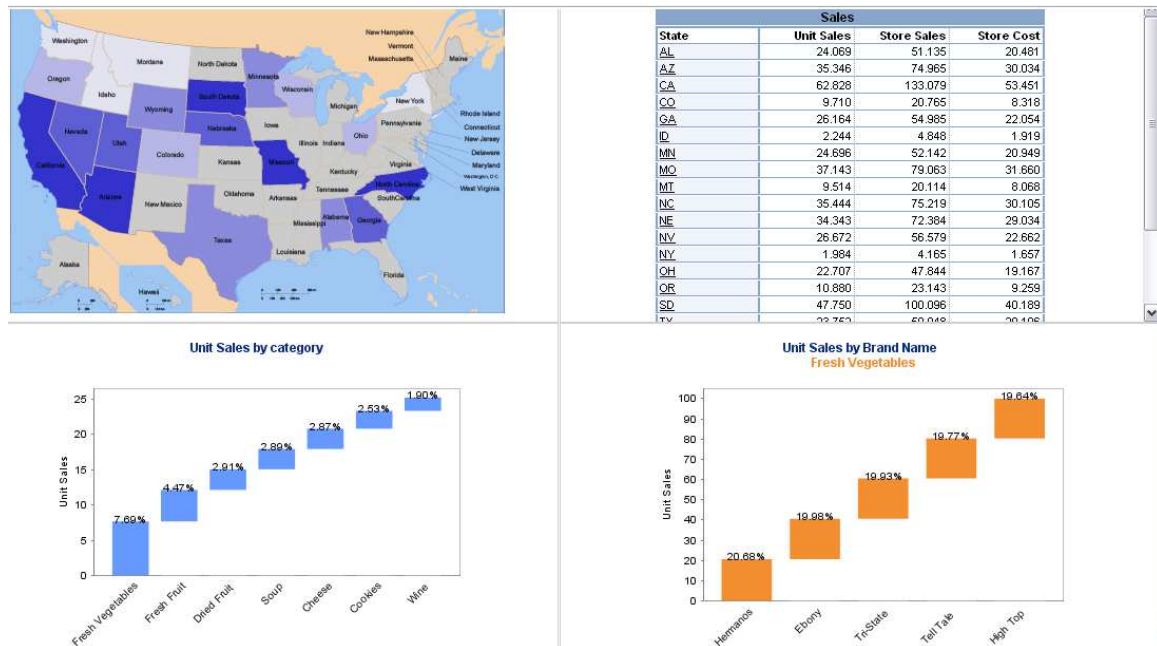


Figura 4.8. Um aspecto da suite SpagoBI (fonte: http://wiki.spagobi.org/xwiki/bin/view/spagobi_server/Document+Composite)

4.1.5. Vanilla

Vanilla é uma suite BI open source suportada pelo consórcio francês BPM-Conseil, com a primeira versão a ser lançada em Junho de 2008 (Vanilla, 2014). Fundada em 2004, a BPM-Conseil possui vasta experiência em consultadoria, e com a sua equipa de programadores iniciou o desenvolvimento do Vanilla em 2007. Mais que uma agregação de ferramentas, o Vanilla é uma plataforma que permite definir a sequência completa do negócio graças à grande diversidade de componentes que possui. Sendo uma suite disponibilizada integralmente em open source, é distribuída sob uma licença MPL (Mozilla Public Licence), e desenvolvida em linguagem Java, sendo o seu logótipo o ilustrado na figura 4.9.



Figura 4.9. Logótipo da Vanilla

A suite facultava acesso a suporte técnico e a um fórum, mantidos pela equipa de desenvolvimento. Suporta Unicode para a base de dados de cliente e um repositório com quatro línguas padrão: Inglês, Francês, Espanhol e Português. É composta por uma grande diversidade de componentes de onde se destacam um Web portal, que é o servidor aplicacional Tomcat Vanilla e possui o Vanilla Web Portal, o motor runtime para objectos

BI integrado no portal Vanilla, os módulos Web para utilizadores finais e também alguns módulos para Web designers. FreeMetadata, Vanilla plugins for BIRT, Free Dashboards, FreeWebReport, Analysis Schema Designer, BiGateway para ETL, BiWorkflow e BiProcessManager são as outras components da suite.

Entre as suas principais características podemos destacar a integração segura de metadados com o BIRT com optimizações SQL, bem como com o FreeDashboards e Cubes. Permite ainda integração com ETL e Workflow. Possui uma base de dados de indicadores de desempenho (KPIs), totalmente integráveis com outros módulos como o BiGateway ou Dashboard. Possibilita ainda realizar todas as etapas padrão de ETL, com suporte a SQL, OLAP, FreeMetrics e origens de dados FreeMetada.

O suporte para SaaS (Software as a Service) permite hospedar diferentes projectos num único servidor, disponibilizando ainda multi-repositórios. A arquitectura de cluster é suportada pelo módulo Hypervision, e permite gerir clusters de diferentes serviços Vanilla (ETL, relatórios, etc.).

A suite é bastante completa e robusta, com a vantagem de ser totalmente open source, e entre as diversas funcionalidades que disponibiliza ainda se destaca o suporte a Workflow BPMN 2.0 integrado para desenho de processos de negócio complexos, bem como inúmeros recursos de BI como colunas dinâmicas, origens alternativas de dados, validação de políticas de segurança e cálculos personalizáveis no FreeMetadata.

Suporta ainda a gestão integrada de documentos, e permite a integração completa de módulos e documentos: Metadata, Dashboard, Report, Cubes, KPI, Workflow e ETL. A versão mais recente do Vanilla é a 4.4.2, lançada em Dezembro de 2013, ilustrando a figura 4.10 um aspecto desta mesma suite.

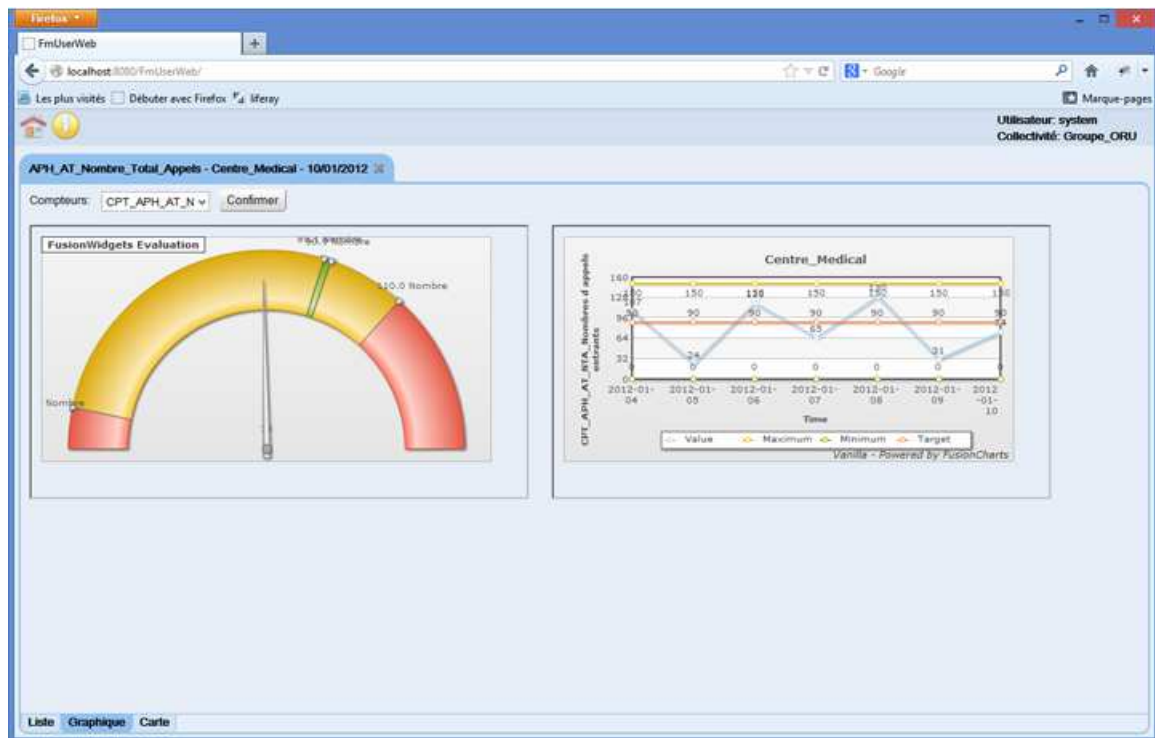


Figura 4.10. Um aspecto da suite Vanilla (fonte: <http://bpm-conseil.com/content/reporting-and-dashboard>)

4.2. Metodologias de avaliação

A avaliação das ferramentas foi baseada nos critérios que a Gartner definiu no seu Magic Quadrant de 2013 (Schlegel et al., 2013). Estes critérios foram por nós adaptados para a realidade das suites open source, e traduzem-se em 11 capacidades divididas por 3 categorias de funcionalidades: i) integração, ii) entrega (disponibilidade) da informação e iii) análise. Os critérios de classificação para cada uma das funcionalidades, são descritos nas secções seguintes.

4.2.1. Integração

- ETL – As ferramentas de ETL permitem efectuar a extracção de dados de diversos sistemas, transformar esses dados segundo as regras de negócio, e efectuar o seu carregamento para data marts ou data warehouses.
- Colaboração – Esta capacidade permite que os utilizadores de BI partilhem e discutam a informação disponibilizada, com chat e notas, e que a possam integrar noutras ferramentas colaborativas ou sociais.

4.2.2. Entrega da informação

- Reporting – Disponibiliza a capacidade de criar reports interactivos formatados, com ou sem parâmetros, com distribuição altamente escalável e capacidades de agendamento. Adicionalmente deve disponibilizar um vasto conjunto de estilos de

reports pré-definidos (por exemplo, financeiros, dashboards operacionais e de desempenho), e deve permitir aos utilizadores o acesso e interacção com os conteúdos disponibilizados em diversas plataformas incluindo Web, dispositivos móveis, e ambientes comuns de portal.

- Dashboards – Este subconjunto de reporting inclui a capacidade de publicar formalmente, baseado em Web ou para dispositivos móveis, reports com mostradores interactivos de informação. Estes mostradores indicam o estado das métricas de desempenho comparadas com um objectivo. São ainda usados para disseminar dados em tempo real de aplicações operacionais ou em conjunção com complexos eventos de processamento.
- Pesquisas ad-hoc – Esta capacidade permite que os utilizadores façam as suas perguntas de dados, sem solicitar a criação de um report. As ferramentas devem possuir uma camada semântica para permitir que os utilizadores naveguem pelas origens de dados.
- Integração com Office – Em muitos casos as plataformas de BI são usadas como uma camada intermédia para gerir, proteger e executar tarefas de BI, sendo o Office (particularmente o Excel) a agir como cliente de BI. Nestes casos, é vital que as ferramentas permitam integração com as aplicações do Office, incluindo suporte para formatos de documentos e apresentações, fórmulas, refrescamento de dados e tabelas dinâmicas. As características “cell locking” e “write-back” serão benefícios adicionais e tidas como integração avançada.
- BI Móvel – Esta capacidade permite que as organizações apresentem reports e dashboards nas plataformas móveis, numa publicação e/ou modo interactivo, e aproveitar o modo de interacção dos dispositivos e outras capacidades não disponíveis normalmente nos monitores.

4.2.3. Análise

- OLAP – Permite que os utilizadores finais analisem dados com pesquisas bastante rápidas e desempenho de cálculo, navegando facilmente em pesquisas multidimensionais. Nalguns casos permite o write-back de valores para bases de dados proprietárias para planeamento e modelação. Pode abranger uma variedade de arquitectura de dados (relacional, multidimensional) e de armazenamento (disco ou memória).
- Visualização interactiva – Oferece ao utilizador a possibilidade de visualizar numerosos aspectos dos dados mais eficientemente, usando imagens e gráficos, em vez de linhas e colunas. Ao longo do tempo, a visualização avançada permitirá fatiar os dados para os incluir mais nos processos orientados a projectos de BI,

permitindo que todos os intervenientes compreendam melhor o fluxo de trabalho através da representação visual.

- Modelo preditivo e data mining – Permite explorar grandes quantidades de dados à procura de padrões consistentes, como regras de associação ou sequências temporais, para detectar relacionamentos entre variáveis, detectando assim novos subconjuntos de dados, usando técnicas matemáticas avançadas. Os programadores de projectos BI podem integrar os modelos facilmente em reports, painéis de análise e processos de negócio.
- Scorecards - Torna as métricas exibidas em painéis de análise em algo mais avançado, aplicando-as num mapa estratégico que alinha os indicadores de desempenho (KPIs) com objectivos estratégicos. As métricas dos “scorecards” podem ser inseridas em reports e informações relacionadas, a fim de realizar análises mais aprofundadas. Implica o uso de metodologias de gestão de desempenho.

4.3. Comparação das Suites

Após o estudo das suites open source BI que identificámos como tendo capacidades nas três categorias de funcionalidades estudadas, foi efectuada uma comparação relativamente às funcionalidades descritas na secção anterior. A Tabela 4.1 apresenta a comparação das suites tendo em conta as capacidades suportadas.

Tabela 4.1 – Comparação das suites open source BI

Funcionalidades	Suites				
	<i>JasperSoft</i>	<i>Palo</i>	<i>Pentaho</i>	<i>SpagoBI</i>	<i>Vanilla</i>
ETL	✓	✓	✓	✓	✓
Colaboração		✓		✓	✓
Reporting	✓	✓	✓	✓	✓
Dashboards	✓	✓	✓	✓	✓
Pesquisas ad-hoc		✓	✓	✓	✓
Integração com Office	✓	✓	✓	✓	
BI Móvel	✓		✓	✓	✓
OLAP	✓	✓	✓	✓	✓
Visualização interactiva	✓	✓	✓	✓	✓
Data mining			✓	✓	✓
Scorecards			✓	✓	✓

Desta análise, podemos destacar como a melhor ferramenta do estudo a SpagoBI, que suporta todas as capacidades por nós seleccionadas, sendo que das suites estudadas, é, a par com a Vanilla, uma suite que possui uma única versão, totalmente open source e gratuita. A SpagoBI disponibiliza ainda outras capacidades além das avaliadas neste estudo, como por exemplo Localização Inteligente (GEO-GIS) ou Master Data Management (para usufruto das capacidades write-back na base de dados) e rivaliza mesmo com a maior parte das versões comerciais das outras suites em estudo.

A Pentaho surge como a segunda ferramenta na nossa análise, visto que suportando basicamente as mesmas capacidades da SpagoBI, possui uma versão comercial, o que consideramos uma desvantagem. É no entanto, das estudadas, a suite eventualmente mais usada, como referido por (Schlegel et al., 2013). Oferece, junto com a SpagoBI uma solução completa que permite atingir os objectivos duma implementação BI numa empresa, em busca dum maior conhecimento do negócio e duma vantagem competitiva. Contudo do estudo efectuado, concluímos que a suite não permite que os utilizadores partilhem e discutam a informação disponibilizada, com chat e notas, e que a possam integrar noutras ferramentas colaborativas.

O facto de não suportar uma integração com Office (Microsoft, OpenOffice, LibreOffice ou outros) traduz-se numa desvantagem para a Vanilla, que, sendo das suites menos conhecidas, mas que nas restantes funcionalidades rivaliza tanto com a SpagoBI como com a Pentaho. É junto com estas uma boa solução de implementação de BI.

Já a Palo e a JasperSoft, podemos considerar que estão num nível inferior às suites já referidas. O facto de não suportarem data mining nem scorecards deixa-as em grande desvantagem relativamente à concorrência, pois negligenciam nas versões open source capacidades fundamentais nas implementações de BI, apostando essencialmente nas versões comerciais, nas quais disponibilizam outro tipo de capacidades. O facto da Palo não disponibilizar uma versão desde Julho de 2011 coloca-a, no nosso entender, para a última posição deste estudo, pois não disponibiliza soluções de mobilidade, tão usadas actualmente, e que são uma tendência do mercado a par com a análise Big Data (não levado em conta neste trabalho). A JasperSoft peca essencialmente pela discrepância de oferta da versão open source relativamente às suas versões comerciais, que são três, apostando visivelmente mais nessa vertente. No entanto, relativamente às capacidades disponibilizadas na versão Community, estas são robustas e competitivas com as restantes.

Resumindo, neste capítulo, descrevemos e analisámos as suites open source escolhidas, e comparámo-las pelas suas funcionalidades. Da comparação escolhemos as duas ferramentas que mais funcionalidades e potencialidades ofereciam e que irão ser alvo de avaliação experimental. Foram estas as suites Pentaho e SpagoBI.

No capítulo seguinte faremos uma avaliação experimental a estas ferramentas, para seleccionar aquela com que iremos realizar a implementação.

5. AVALIAÇÃO EXPERIMENTAL

Após realizar a análise das ferramentas baseado na informação disponibilizada por cada fornecedor, neste capítulo irá ser feita a análise prática das duas ferramentas que mais se destacaram nesse estudo: SpagoBI e Pentaho.

A análise será efectuada sobre duas perspectivas, a da avaliação do software open source em si, e das funcionalidades do BI, que serão as já referidas na secção anterior e baseadas nos critérios da Gartner.

Com vista a garantir a qualidade, surgiram diversas metodologias de avaliação de software open source nos últimos anos, que devem ser levadas em conta antes de se tomar a decisão de adopção de um destes tipos de software por parte das organizações. Duas destas metodologias que se destacam são a Qualification and Selection Open Source (QSOS, 2013) apoiado pela Atos Origin (Atos, 2013) e a Business Readiness Rating (OpenBRR, 2013) criada pelo Carnegie Mellon Silicon Valley Center for Open Source Investigation (CMU, 2013) e Intel (Intel, 2013). Ambas as metodologias de avaliação ajudam os utilizadores a alcançar um objectivo semelhante, que é seleccionar entre uma lista de projectos semelhantes o mais adequado ao seu contexto.

Em (Deprez et al., 2008) são descritas as vantagens e desvantagens de cada uma das metodologias, sendo referido que a OpenBRR permite seleccionar os critérios para os adaptar a um contexto, e que a QSOS é ambígua em mais de metade dos seus critérios. Esta análise levou-nos a optar pela adopção da metodologia OpenBRR para avaliar as suites open source BI.

5.1. OpenBRR

A metodologia OpenBRR assenta em 12 categorias que estão descritas na tabela 5.1.

Tabela 5.1 – Categorias da metodologia OpenBRR

Categoria	Descrição
Funcionalidade	O software atende às exigências do utilizador médio?
Usabilidade	O interface é agradável? É de fácil utilização para o utilizador final? É fácil de instalar, configurar, implementar e manter?
Qualidade	Qual a qualidade do desenho de software, código e testes? Quão completa e livre de erros é?
Segurança	Como lida o software com questões de segurança? É seguro?
Desempenho	Como é o desempenho do software?
Escalabilidade	Como responde num ambiente de grande escala?
Arquitectura	Como é a arquitectura do software? É modular, portátil, flexível, extensível, aberto e fácil de integrar?
Apoio	Como é a componente de apoio do software?

Documentação	Qual a qualidade da documentação do software?
Adopção	Como é a adopção do software pela comunidade, mercado e indústria?
Comunidade	A comunidade de suporte do software está activa e animada?
Profissionalismo	Qual o nível de profissionalismo do processo de desenvolvimento e da organização do projecto como um todo?

Neste modelo, a classificação das categorias pode ter diferentes níveis de importância dependendo dos requisitos para utilização do software. A implementação do método divide-se em quatro fases.

1. Numa primeira fase deste método, deve-se identificar a lista de componentes do software a ser analisada, medir cada componente em relação aos critérios de avaliação, e remover da análise qualquer componente do software que não satisfaça os requisitos de uso.
2. Na segunda fase, deve-se atribuir pesos às categorias e às métricas do seguinte modo:
 - Classificar as 12 categorias de acordo com a importância (1 – a mais alta, 12 – a menor).
 - Seleccionar as 7 categorias mais altas (ou menos), e atribuir uma percentagem de importância para cada uma, totalizando 100%.
 - Para cada métrica dentro de uma categoria, classificar a métrica de acordo com a importância.
 - Para cada métrica dentro de uma categoria, atribuir a importância percentual, totalizando todas as métricas 100% da categoria.
3. Na terceira fase, deve-se recolher a informação para cada métrica utilizada em cada categoria, e atribuir uma ponderação, num nível de 1 (inaceitável) a 5 (excelente).
4. Por fim, na quarta fase, deve-se utilizar a qualificação da categoria e os factores de ponderação para calcular a pontuação final.

A figura 5.1. abaixo mostra de uma forma geral o funcionamento do método.

Business Readiness Rating Model

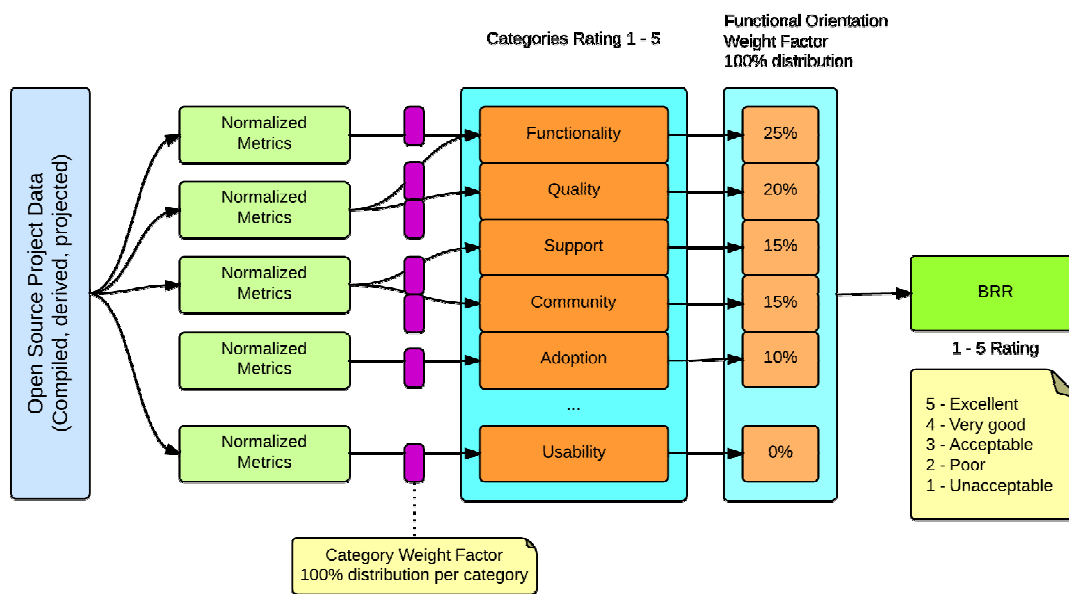


Figura 5.1. Esquema do modelo OpenBRR (baseado na fonte: <http://ecarrerasg.wordpress.com/2012/01/02/openbrr-and-qsos/>)

As métricas por cada categoria da metodologia OpenBRR são mostradas e descritas na tabela 5.2.

Tabela 5.2 – Métricas para cada categoria da metodologia OpenBRR

Categoria	Métrica	Descrição
Usabilidade	Experiência do interface para o utilizador final	Mede o quanto o interface é perceptível ao utilizador final (interface intuitivo, navegação, esquema de controle)
	Tempo de instalação dos pré-requisitos para a instalação do software open source	Tempo/esforço necessário para colocar o sistema funcional, com todos os pré-requisitos instalados. Não inclui o sistema operativo
	Tempo para instalação/configuração	Tempo para obter a <i>instant gratification</i> , mostra se o projecto está pensado para levar os utilizadores a funcionar o mais rapidamente possível
Qualidade	Nº de <i>minor releases</i> nos últimos 12 meses	Mede os updates planeados e a correcção de erros. Normalmente, os <i>service packs</i> nos produtos

		comerciais
	Nº de <i>point/patch releases</i> nos últimos 12 meses	Normalmente são correcções para <i>P1 bugs</i> ¹ como deadlocks, memória e vulnerabilidades de segurança
	Nº de <i>bugs</i> relatado nos últimos 6 meses	Mede a qualidade do uso do projecto
	Nº de <i>bugs</i> resolvidos nos últimos 6 meses (comparado com o nº de <i>bugs</i> relatados)	Mede a rapidez de resolução de <i>bugs</i>
	Nº de <i>P1/critical bugs</i> relatados	Mede a gravidade dos problemas de qualidade encontrados
	Tempo médio de <i>P1 bugs</i> nos últimos 6 meses	Mede a capacidade de resposta a resolução de problemas críticos
Segurança	Número de vulnerabilidades de segurança nos últimos seis meses que são moderadas a extremamente críticas	Mede a qualidade relativamente a vulnerabilidades de segurança. Quão susceptível está o software a vulnerabilidades de segurança
	Número de vulnerabilidades de segurança ainda em aberto (sem correcção)	Mede a capacidade que o projecto tem em resolver todas as vulnerabilidades de segurança
	Há informação dedicada para segurança (página web, wiki, etc)?	Mede o conhecimento e quanto são levados a sério os problemas de segurança
Desempenho	Há testes de desempenho e relatórios de referência (Benchmark) disponíveis?	Mede se há testes de desempenho e benchmarks publicados – tipicamente em comparação com soluções equivalentes
	Optimização de desempenho e configuração	Mede se há alguma documentação ou ferramenta que ajude a afinar o desempenho (informação sobre CPU, disco, rede...)
Escalabilidade	Referência de implementação	Mede o quanto o software é escalável e testado em ambiente real, através de implementação no mundo real
	Projectado para escalabilidade	Mede se o projecto foi concebido com a escalabilidade em mente. É <i>thread-safe</i> ? Pode ser executado

¹ *P1 bug* é um bug de prioridade 1 que normalmente significa que sem a sua resolução a aplicação não pode continuar

		em ambiente de <i>cluster</i> ? Pode o hardware resolver problemas de desempenho?
Arquitectura	Há <i>plug-ins</i> de terceiros?	Mede se o projecto foi desenhado para extensibilidade por terceiros
	API pública / Serviço Externo	Permite extensões via API pública, também permite personalização?
Apoio	Volume médio de listas de discussão gerais nos últimos 6 meses	As listas de discussão gerais são o primeiro sítio onde as pessoas vão para ajuda gratuita
	Qualidade do apoio profissional	Apoio profissional que ajuda a afinar a implementação e resolução de problemas é sempre desejável
Documentação	Existência de vários tipos de documentação	Uma boa ferramenta de documentação deve incluir documentação para vários grupos de utilizadores em vários formatos
	Contribuição do utilizador	As melhores ajudas vêm muitas vezes de exemplos fornecidos por utilizadores. É o feedback de pessoas que usaram o produto
Adopção	Quantos títulos de livros na Amazon.com devolve a consulta: "subject:computer and title:component name"	A disponibilidade de livros é certamente bom. Uma forma padronizada de contagem dos livros disponíveis é importante
	Referência de implementação	Mede o quanto o software é escalável e testado em ambiente real, através de implementação no mundo real
Comunidade	Volume médio de listas de discussão geral nos últimos 6 meses	As listas de discussão gerais são o local onde a comunidade se interajuda
	Nº de contribuintes de código nos últimos 6 meses	Os contribuintes de código normalmente promovem a construção de uma comunidade à volta do projecto. Quanto maior o número de contribuintes de código, melhor o apoio da comunidade
Profissionalismo	<i>Driver</i> do projecto	O <i>driver</i> do projecto realiza a gestão de projectos, angariação de dinheiro, etc.

	Dificuldade de pertencer à equipa de desenvolvimento	Para garantir a qualidade do software, projectos maduros devem ser selectivos na aceitação de membros. Novos projectos não têm opção
--	--	--

A categoria “Funcionalidade” é avaliada de forma diferente das outras, e segue um método próprio. Nesta categoria pretende-se analisar e avaliar as características que as ferramentas têm ou deveriam ter (aquilo que se considere ser expectável), mais extras. No caso das suites open source BI o ETL, colaboração, reporting, etc. O método para avaliação desta categoria é o seguinte:

1. Definir os recursos (características) a analisar, atribuindo-lhes uma importância de 1 a 3 (menos importante a muito importante);
2. Classificá-las numa soma cumulativa (de 1 a 3), podendo atribuir-se nota negativa (retirar pontos à soma cumulativa) se não cumprirem uma característica;
3. Fazer o mesmo para recursos extra;
4. Dividir a soma cumulativa (incluindo extras) pela pontuação máxima possível de obter pelas características base (sem extras). Isto intencionalmente pode dar inferior ou igual a 0 ou superior a 100, valorizando os extras, e desvalorizando (punindo) as características básicas não contempladas;
5. Normalizar o resultado anterior para uma escala de 1 a 5 para a categoria Funcionalidade, na seguinte escala:
 - Menos de 65%, nota = 1 (inaceitável)
 - 65% - 80%, nota = 2 (fraco)
 - 80% - 90%, nota = 3 (aceitável)
 - 90% - 96%, nota = 4 (bom)
 - Superior a 96%, nota = 5 (excelente)

Esta escala final enquadra-se então nas mesmas medições do resto das características. Estas métricas são um conjunto representativo do modelo Open BRR, entendendo-se que poderá haver melhores candidatos para métricas para medir a prontidão de negócio em cada categoria. Para algumas métricas a informação é difícil de recolher, ou está simplesmente indisponível.

5.2. Avaliação

O método Open BRR, estabelece que se devem escolher, no máximo, as 7 categorias mais relevantes para a análise, de forma a evitar que as mais importantes tenham a sua importância diluída pelas menos relevantes. Após esta escolha, deve então atribuir-se-lhes o peso para a avaliação final.

A tabela 5.3 mostra a escolha feita para esta avaliação, tendo em conta os interesses da implementação.

Tabela 5.3 – Peso das diversas categorias escolhidas

Categoria	Peso
Funcionalidade	40%
Usabilidade	15%
Qualidade	15%
Documentação	10%
Comunidade	10%
Segurança	5%
Arquitetura	5%

A justificação para esta escolha é que as características desempenho, escalabilidade, apoio, adopção e profissionalismo são pouco relevantes relativamente às escolhidas tendo em vista o tipo de implementação, a dimensão da empresa, a dimensão do seu departamento informático, e a dimensão das próprias ferramentas em análise.

Sendo uma implementação prática numa empresa, com o propósito de lhe trazer uma clara vantagem competitiva, a categoria funcionalidade foi aquela que teve um peso maior na distribuição. A usabilidade e a qualidade têm a ordem de importâncias seguintes pois tratam-se daquelas que vão causar impacto no utilizador final. A documentação e a comunidade seguem-se pelo facto de serem importantes para o departamento de TI da empresa poder realizar a manutenção, continuidade e evolução após a implementação. Sendo importantes, mas devido às métricas sugeridas pelo método para a sua avaliação, a segurança e a arquitectura foram consideradas mas coube-lhes o menor peso na análise.

A categoria desempenho não foi considerada pelo facto de não haver benchmarks publicados sobre estas ferramentas, e por estas não disponibilizarem praticamente nenhuma informação sobre a melhor forma de o otimizar. Sendo o software em análise suites que têm muitas implementações e estão em constante evolução, a escalabilidade também não foi considerada. Como o apoio se refere em parte a apoio profissional, e pelo facto de já ser prática na empresa o uso de software open source, com interacção com comunidades de programadores, e ainda por ter sido considerada para análise a categoria comunidade, esta categoria também foi considerada como sendo das menos relevantes. A adopção também foi considerada pouco relevante, pois ambas as suites têm forte implementação no mercado, como já referido. Por fim, e como estas suites têm comunidades bastante relevantes, e pelo facto de serem ambas suportadas num estilo corporativo, com grandes empresas a suportá-las, também o profissionalismo foi uma categoria considerada das menos relevantes.

Para avaliar a categoria da funcionalidade, foram consideradas as capacidades já usadas para a escolha das duas ferramentas, a que se atribuiu a sua relevância na escala de 1 (pouco importante) a 3 (muito importante), como mostra a tabela 5.4.

Tabela 5.4 – Peso das capacidades da categoria funcionalidades

Capacidades	Peso
ETL	3
Colaboração	2
Reporting	3
Dashboards	3
Pesquisas ad-hoc	2
Integração com Office	2
BI Móvel	2
OLAP	3
Visualização interactiva	1
Data mining	1
Scorecards	2

Pela importância para a empresa, na utilização final, foi atribuído o maior peso (3) ao ETL, Reporting, Dashboards e OLAP, importância intermédia (2) à Colaboração, Pesquisas ad-hoc, Integração com Office, BI Móvel e Scorecards, e menor importância (1) à Visualização interactiva e ao Data mining.

A avaliação das funcionalidades das suites, é apresentada na tabela 5.5 abaixo.

Tabela 5.5 – Avaliação das funcionalidades das suites

Funcionalidades esperadas	Peso	Classificação	
		Pentaho	SpagoBI
ETL	3	3	3
Colaboração	2	2	2
Reporting	3	3	3
Dashboards	3	3	3
Pesquisas ad-hoc	2	2	2
Integração com Office	2	0	2
BI Móvel	2	2	2
OLAP	3	3	3
Visualização interactiva	1	1	1
Data mining	1	1	1
Scorecards	2	2	2
Funcionalidades Extra			
Georeferenciação	2	0	2
Totais	26	22	26

A funcionalidade de Georeferenciação foi incluída na análise por ter sido demonstrado interesse por parte da empresa na sua inclusão, pois sendo atractiva visualmente para o utilizador final, pode ajudar a quebrar barreiras sobre a forma como os utilizadores finais

olham geralmente para mudanças tecnológicas. Tendo a SpagoBI essa funcionalidade, constitui uma vantagem para esta suite nesta avaliação. Às funcionalidades não suportadas, atribuímos a nota 0.

Na avaliação das Funcionalidades, segundo a escala do Open BRR, a Pentaho terá nota 3, com 84,6% (22 pontos em 26), e a SpagoBI nota 5, com 100% (26 pontos em 26).

Para as restantes categorias, Usabilidade, Qualidade, Documentação, Comunidade, Segurança e Arquitectura, foi utilizado o critério estabelecido pelo método Open BRR, como mostra a avaliação nas tabelas 5.6 a 5.11, abaixo.

O critério para esta avaliação foi o de análise da informação fornecida por cada ferramenta nas páginas Web de cada uma, nos seus fóruns, e nas ferramentas de integração contínua onde podem ser avaliados diversos critérios, como o nº de *bugs* e de contribuintes de código.

A avaliação métrica a métrica, que justifica os valores da avaliação de cada categoria, pode ser encontrada nos anexos, e todos os dados reportam à data da avaliação realizada em Maio de 2013.

Na tabela 5.6 podemos observar a avaliação para a categoria Usabilidade.

Tabela 5.6 – Avaliação da categoria Usabilidade

Categoria		Peso	Nota sem peso		
			Pentaho	SpagoBI	
Usabilidade		15%	5	5	
Métrica	Peso	Nota sem peso		Nota Ponderada	
		Pentaho	SpagoBI	Pentaho	SpagoBI
Experiência do interface para o utilizador final	50%	5	5	2,5	2,5
Tempo de instalação dos pré-requisitos para a instalação do software open source	25%	5	5	1,25	1,25
Tempo para instalação/configuração	25%	5	5	1,25	1,25

Em todas as métricas foi dada nota máxima a ambas as suites, pois em ambas a experiência de utilização é excelente, e o tempo de instalação dos pré-requisitos e da própria suite é menor do que 10 minutos. A nota de ambas as suites na categoria usabilidade foi de 5 pontos, nos 5 possíveis.

Na tabela 5.7 podemos observar a avaliação para a categoria Qualidade.

Tabela 5.7 – Avaliação da categoria Qualidade

Categoria		Peso	Nota sem peso		
			Pentaho	SpagoBI	
Qualidade		15%	1,9	3,4	
Métrica	Peso	Nota sem peso		Nota Ponderada	
		Pentaho	SpagoBI	Pentaho	SpagoBI
Nº de <i>minor releases</i> nos últimos 12 meses	20%	3	5	0,6	1
Nº de <i>point/patch releases</i> nos últimos 12 meses	20%	1	3	0,2	0,6
Nº de <i>bugs</i> relatado nos últimos 6 meses	10%	3	4	0,3	0,4
Nº de <i>bugs</i> resolvidos nos últimos 6 meses (comparado com o nº de <i>bugs</i> relatados)	10%	4	2	0,4	0,2
Nº de <i>P1/critical bugs</i> relatados	20%	1	1	0,2	0,2
Tempo médio de <i>bugs P1</i> nos últimos 6 meses	20%	1	5	0,2	1

Nesta categoria encontramos para cada métrica resultados bem distintos entre as suites, em que, por exemplo, a Pentaho lançou uma versão da suite e a SpagoBI lançou duas, tendo a Pentaho uma nota pior, como pode ser consultado nos anexos. No total da categoria a Pentaho teve uma nota de 1.9 e a SpagoBI teve uma nota de 3.4 em 5 pontos possíveis, oferecendo a SpagoBI maior qualidade segundo estes critérios.

Na tabela 5.8 podemos observar a avaliação para a categoria Documentação.

Tabela 5.8 – Avaliação da categoria Documentação

Categoria		Peso	Nota sem peso		
			Pentaho	SpagoBI	
Documentação		10%	5	5	
Métrica	Peso	Nota sem peso		Nota Ponderada	
		Pentaho	SpagoBI	Pentaho	SpagoBI
Existência de vários tipos de documentação	50%	5	5	2,5	2,5
Contribuição do utilizador	50%	5	5	2,5	2,5

Na categoria de Documentação ambas obtiveram a nota máxima sendo a nota final de 5 pontos em 5.

Na tabela 5.9 podemos observar a avaliação para a categoria Comunidade.

Tabela 5.9 – Avaliação da categoria Comunidade

Categoria		Peso	Nota sem peso		
			Pentaho	SpagoBI	
Comunidade		10%	3,8	3,8	
Métrica	Peso	Nota sem peso		Nota Ponderada	
		Pentaho	SpagoBI	Pentaho	SpagoBI
Volume médio de listas de discussão geral nos últimos 6 meses	60%	3	3	1,8	1,8
Nº de contribuintes de código nos últimos 6 meses	40%	5	5	2	2

Nesta categoria, ambas estão dentro dos mesmos intervalos de avaliação das métricas, pelo que obtiveram a mesma nota final de 3.8 pontos em 5.

Na tabela 5.10 podemos observar a avaliação para a categoria Segurança.

Tabela 5.10 – Avaliação da categoria Segurança

Categoria		Peso	Nota sem peso		
			Pentaho	SpagoBI	
Segurança		5%	5	5	
Métrica	Peso	Nota sem peso		Nota Ponderada	
		Pentaho	SpagoBI	Pentaho	SpagoBI
Número de vulnerabilidades de segurança nos últimos seis meses que são moderadamente a extremamente críticas	40%	5	5	2	2
Número de vulnerabilidades de segurança ainda em abertos (sem correcção)	40%	5	5	2	2
Há informação dedicada para segurança (página web, wiki, etc)?	20%	5	5	1	1

Na categoria de segurança ambas as suites obtiveram nota máxima em todas as métricas, tendo obtido cada uma a nota final na categoria de 5 pontos em 5 possíveis.

Na tabela 5.11 podemos observar a avaliação para a categoria Arquitectura.

Tabela 5.11 – Avaliação da categoria Arquitectura

Categoria		Peso	Nota sem peso		
			Pentaho	SpagoBI	
Arquitectura		5%	5	3	
Métrica	Peso	Nota sem peso		Nota Ponderada	
		Pentaho	SpagoBI	Pentaho	SpagoBI
Há <i>plug-ins</i> de terceiros?	50%	5	1	2,5	0,5
API pública / Serviço Externo	50%	5	5	2,5	2,5

Na categoria Arquitectura a Pentaho destacou-se com os *plug-ins* de terceiros que existem para a suite, em oposição à sua ausência para a SpagoBI. A nota final da categoria foi de 5 pontos em 5 para a Pentaho e de 3 em 5 para a SpagoBI.

Ponderando as notas de cada categoria, obtivemos o resultado da tabela 5.12, abaixo.

Tabela 5.12 – Nota ponderada das suites em avaliação

Categoria	Peso	Nota Ponderada	
		Pentaho	SpagoBI
Funcionalidade	40%	1,2	2
Usabilidade	15%	0,75	0,75
Qualidade	15%	0,285	0,51
Documentação	10%	0,5	0,5
Comunidade	10%	0,38	0,38
Segurança	5%	0,25	0,25
Arquitectura	5%	0,25	0,15
		3,615	4,54

Aplicando o método Open BRR, com a escolha das categorias mais relevantes tendo em conta a implementação real e as necessidades da empresa, e distribuídos os pesos tendo em conta o mesmo interesse, a SpagoBI atinge a nota mais alta entre as duas suites.

No critério do método Open BRR, a nota de 4,54 é “Muito Bom”, o que oferece toda a confiança para ser esta a suite escolhida para prosseguir o trabalho.

A aplicação deste método mostrou-se útil nesta avaliação, pois oferece um critério de avaliação objectivo, reproduzível e auditável, atribuindo um valor numérico que permite uma comparação com outras avaliações de outros produtos.

Assim, será a SpagoBI que será utilizada na implementação em ambiente real na empresa.

Em resumo, neste capítulo descrevemos a metodologia de avaliação e avaliámos as duas suites que escolhemos anteriormente, a SpagoBI e a Pentaho. Desta avaliação resultou a suite com que faremos a implementação.

No capítulo seguinte, faremos a análise ao modelo de dados da data warehouse, tendo em vista a implementação que responderá às necessidades da empresa.

6. MODELO DE DADOS

Neste capítulo será descrita a empresa MICROplásticos, SA, onde será realizada a implementação da suite SpagoBI, e analisaremos as necessidades a que teremos que dar resposta. Será também definido o modelo dimensional que servirá para construir a data warehouse que suportará a implementação.

6.1. A empresa

A MICROplásticos, S.A., sediada na Figueira da Foz, iniciou a sua actividade em 1987 e é uma empresa que desenvolve uma actividade industrial de fabrico de peças plásticas de alta precisão e de componentes funcionais para a indústria eléctrica, electrónica, de consumo e automóvel.

A actividade principal é a injeção de peças em plástico - transformação de matérias plásticas para produção de componentes industriais. Emprega, em média, cerca de 230 trabalhadores e tem influência directa sobre o emprego de mais 150 trabalhadores ligados a pequenas empresas locais prestadoras de serviços. Não produz directamente para venda ao consumidor final. Os clientes são pertencentes a grupos transnacionais e fornecedores de grandes indústrias. O volume de negócio anual foi de 27 milhões de euros (Diário Coimbra, 2014) para o ano de 2012.

A produção está centrada em quatro indústrias – produtos de consumo, eléctrica, electrónica e automóvel, sendo que esta última é a que tem mais peso. No entanto, a MICROplásticos não está refém de nenhuma das actividades deste sector, nem esta predominância determina qualquer perda de competitividade para responder a solicitações de outros clusters, distribuindo as vendas por várias áreas dentro da indústria automóvel (v.g. segurança, electrónica, interiores).

É uma empresa predominantemente exportadora. O destino maioritário das suas vendas é o mercado espanhol, em consonância com a prática da economia nacional, pois Espanha tem um posicionamento bastante forte na indústria automóvel, ocupando o terceiro lugar como produtora de veículos e componentes e é uma referência neste cluster em termos de economia global.

MICROplásticos é detentora de certificação dos Sistemas de Gestão Ambiental (ISO 14000), do Sistema de Gestão de Segurança e Medicina (OSHAS 18001/NP 4397), do Sistema da Qualidade (NP EN ISO 9001:2000 e ISO/TS 16949:2002).

O processo de fabrico tem como principal ferramenta o molde, construído em aço, que possui cavidades com a forma da peça a produzir. Cada molde serve apenas para produzir uma peça ou conjunto de peças e é propriedade do cliente, pois a peça plástica que daí resulta é exclusiva. A matéria-prima é plástico que tem um aspecto granulado. Durante o processo de produção sofre um aquecimento (processo de estufagem) à entrada na

máquina de injeção para o derreter e depois é injectada no molde onde adquire a forma aí desenhada.

A figura 6.1. mostra o organograma da empresa.

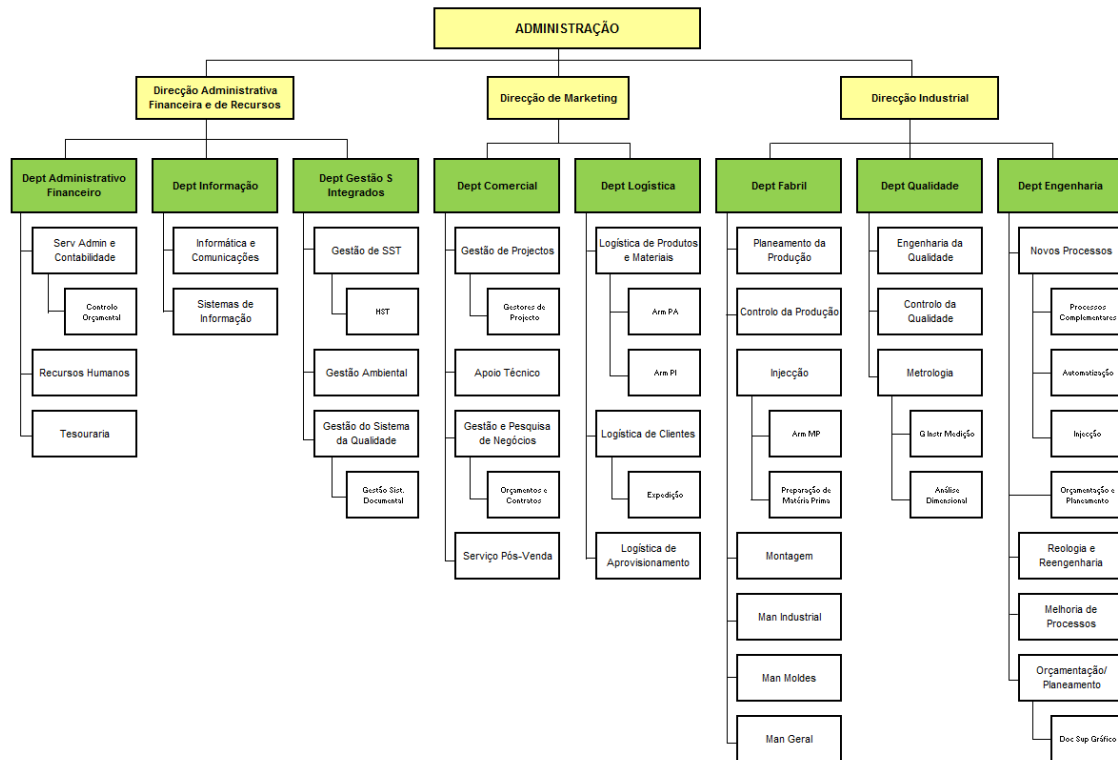


Figura 6.1. Organograma da empresa (fonte: MICROplásticos, SA)

Nesta secção apresentámos a empresa onde será realizada a implementação. Na secção seguinte analisaremos as questões a que a implementação do BI na empresa deve responder.

6.2. Questões a responder

No âmbito desta implementação, foi decidido com a MICROplásticos actuar e responder a questões sobre quatro áreas que actualmente têm deficiência de informação e bastante trabalho para obterem os indicadores necessários ao seu negócio e ao seu processo de decisão: Recursos Humanos, Administração, Departamento Comercial e Logística.

Para cada uma destas áreas as questões que se pretendem ver respondidas são as que estão descritas nas secções seguintes.

6.2.1. Recursos Humanos

Os Recursos Humanos pretendem ver respondidas as seguintes questões:

- Nº pessoas por tipo de horário²/ turno
- Absentismo por departamento/ secção³ (Análise por dia da semana/ mês/ turno)
- Absentismo por tipo de falta (Análise por dia da semana/ mês/ turno)
- Horas trabalhadas

6.2.2. Administração

A Administração elegeu como principais questões a serem respondidas as seguintes:

- Gastos com Pessoal/ Vendas
- Margem bruta⁴ das vendas de Produto Acabado (PA)
- Margem bruta Total
- Margem bruta dos Moldes
- Peso dos consumos⁵ sobre as vendas
- Distribuição dos consumos por tipo (MP Plást, Comp Plást, Comp. Diversos, Embalagem)
- Distribuição dos FSE⁶
- Subcontratação - Injecção
- Subcontratação - Montagem
- VAB⁷/ Nº colaboradores
- VAB/ Gastos com pessoal

6.2.3. Comercial

O departamento Comercial pretende obter resposta para as seguintes questões:

- Vendas por tipo de operação (Injecção; Injecção + Montagem; Montagem)
- Vendas de Produto Acabado (PA)
- Vendas de PA - real vs objectivo
- Vendas de PA - por cliente
- Vendas de PA - por mercado geográfico
- Vendas de PA - por indústria

² Há funcionários que trabalham por turnos e outros por horário normal

³ Uma secção implica logo um departamento

⁴ Margem Bruta = Vendas / Custo da matéria prima

⁵ Consumo = Custo da matéria prima

⁶ FSE – Fornecimentos e serviços externos

⁷ VAB - Valor acrescentado bruto = Vendas – Custo mercadorias vendidas (61) – Gastos com pessoal (63)

- Vendas de projectos
- Vendas de projectos - por cliente
- Vendas de projectos - por mercado geográfico
- Vendas de projectos - por indústria

6.2.4. Logística

A logística elegeu como questões a ver respondidas as seguintes:

- Variação da produção⁸
- Cobertura de stocks⁹ Produto Acabado (PA)
- Cobertura de stocks Matéria Prima (MP)
- Transportes especiais/ Venda de produtos
- Transportes especiais
- Transportes por tipo

Na secção seguinte será definido o modelo dimensional da data warehouse a implementar por forma a responder às questões que os diversos departamentos pretendem obter resposta.

6.3. Modelo dimensional

Nesta secção, tendo em conta as questões a responder, analisaremos o modelo dimensional para implementar a data warehouse.

Para responder às diversas questões do modelo de negócio da MICROplásticos, a granularidade definida foi o dia. Não se justifica neste caso um detalhe maior, e é suficientemente detalhado para responder às diversas necessidades.

Assim, foram identificadas seis tabelas de factos divididas pelas quatro áreas de negócio abrangidas.

Nos Recursos Humanos, como tabelas de factos, temos

- Laboração – factos referentes à laboração. Por data, por funcionário, por área funcional e por turno. Dá-nos informação sobre o registo de ponto, nomeadamente a Hora de Entrada, Hora de saída, nº de horas trabalhadas, de falta, e extra.
- Absentismo – factos referentes ao absentismo. Por data, por funcionário, por área funcional, por turno e por tipo de falta, dá-nos o nº de horas e de minutos das faltas.

⁸ Stock final – inicial, por período (sempre em valor)

⁹ Stock no fim do mês, multiplicado por 30, sobre as vendas do mês (sempre em valor)

A inclusão das chaves forasteiras por área funcional e por turno nestas duas tabelas, justificam-se pelo facto de ocasionalmente haver rotatividade de funcionários entre áreas funcionais e turnos.

Na área de Administração, temos apenas a tabela de factos:

- Contabilidade – factos referentes aos movimentos contabilísticos. Por data e por conta do SNC (sistema de normalização contabilística), dá-nos informação sobre os valores a débito, a crédito e saldo (débito menos crédito) de determinada conta do SNC.

Na área Comercial, temos a tabela de factos:

- Vendas – factos relativos a vendas. Por data, por cliente/fornecedor e por produto, dá-nos se foi uma venda de injeção e/ou montagem, a sua quantidade e valor.

Utiliza-se cliente/fornecedor em vez de apenas cliente, uma vez que determinadas entidades são em simultâneo uma coisa e outra, isto é, uma entidade pode fornecer matéria-prima e comprar peças.

Por fim, na área de Logística, temos as tabelas de factos:

- Stock – factos relativos a stock de produtos. Por data, por produto e por armazém, dá-nos a quantidade, o preço unitário e valor do stock.
- Transportes – factos relativos a transportes. Por data, por transportadora, por tipo, e por natureza, dá-nos informação se foi transporte normal ou especial, se foi produto acabado ou matéria-prima, interno ou externo, se foi subcontratação, se foi da categoria diversos, e qual foi o seu valor de custo.

A valorização do stock de produto acabado e de matéria-prima é diferente, sendo respectivamente por custo padrão e por custo médio ponderado. No entanto tem sentido manter o stock destes dois tipos de produtos, resolvendo-se o seu problema de valorização aquando do ETL.

Na tabela de transportes, o tipo é um código interno da empresa, e identifica se determinado transporte é normal, especial, se foi produto acabado ou matéria prima, mercado interno ou externo, se foi subcontratação ou diversos, segundo o seguinte código:

- 01 - Normal, Produto Acabado, Mercado Externo.
- 02 - Normal, Produto Acabado, Mercado Interno.
- 03 - Normal, Matéria Prima, Mercado Externo.
- 04 - Normal, Matéria Prima, Mercado Interno.
- 05 - Normal, Subcontratação .
- 06 - Diversos.
- 07 - Especial, Produto Acabado, Mercado Externo.
- 08 - Especial, Produto Acabado, Mercado Interno.
- 09 - Especial, Matéria Prima, Mercado Externo.

- 10 - Especial, Matéria Prima, Mercado Interno.
- 11 - Especial, Subcontratação.

Por sua vez, a natureza indica se foi um débito ou um crédito, isto é se foi um custo para a empresa ou se foi um transporte prestado pela empresa.

Como tabelas de dimensão foram definidas as seguintes:

- Data – Além de toda a informação relativa à data, inclui o nº de funcionários a cada momento visto haver um indicador que é o valor acrescentado bruto a dividir pelo nº de funcionários, tornando necessário saber esse dado, carregada com valores desde o início de 2012.
- Produto – informação sobre cada produto.
- Clientes/Fornecedores – informação sobre os clientes e fornecedores.
- País – informação sobre os países de clientes/fornecedores e ou moradas.
- Funcionário – informação relativa ao funcionário.
- Turno – Informação relativa a cada turno existente – há turnos rotativos, outros que não o são, e há turnos de horário normal.
- Tipo de falta – informação relativa ao tipo de falta.
- Área Funcional – Informação relativa à área funcional da empresa – há rotatividade de funcionários entre áreas funcionais.
- SNC – Informação relativa às contas do sistema de normalização contabilística.
- Armazém – Informação relativa aos armazéns.

Na figura 6.2 podemos observar uma visão genérica do modelo dimensional. Podemos observar que existem quatro áreas de intervenção identificadas, num modelo em estrela, com a granularidade a ser definida pela tabela “Data”, comum a todas as tabelas de factos. Nas figuras seguintes, 6.3 a 6.6, podemos observar detalhes do modelo, por área de actuação.

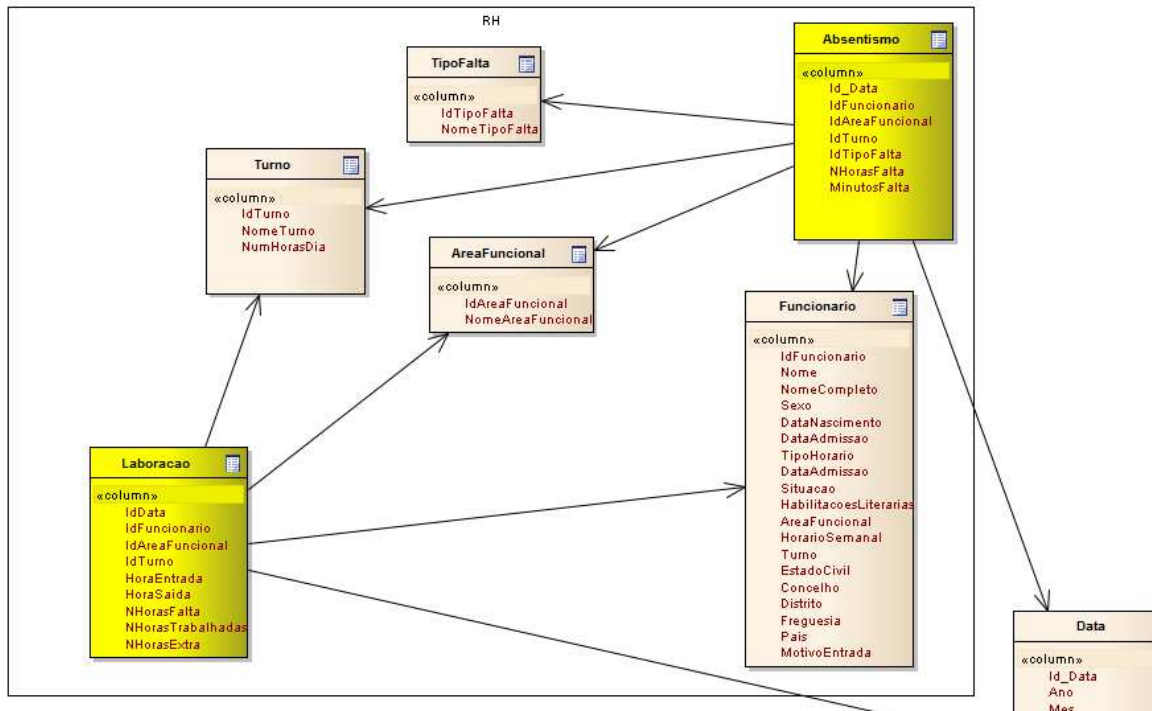


Figura 6.3. Detalhe do modelo dimensional para a área dos Recursos Humanos

Com o detalhe da figura 6.3 conseguimos responder a todas as questões que o departamento dos Recursos Humanos pretende.

Na figura 6.4, podemos observar um detalhe do modelo dimensional para a área da Administração.

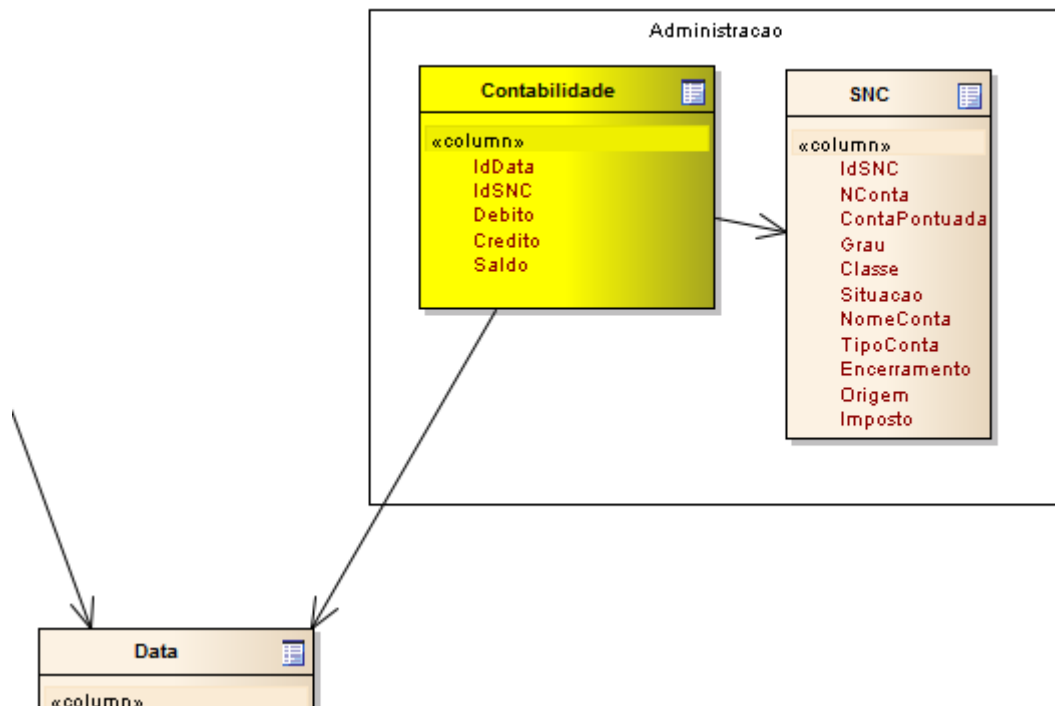


Figura 6.4. Detalhe do modelo dimensional para a área da Administração

A figura 6.4 mostra a tabela de factos “Contabilidade” que tem como dimensões a tabela “SNC” e “Data”, que define a granularidade, e responde a todas as questões que a Administração pretende, baseado no modelo em árvore do SNC.

Na figura 6.5, podemos observar um detalhe do modelo dimensional para a área Comercial.

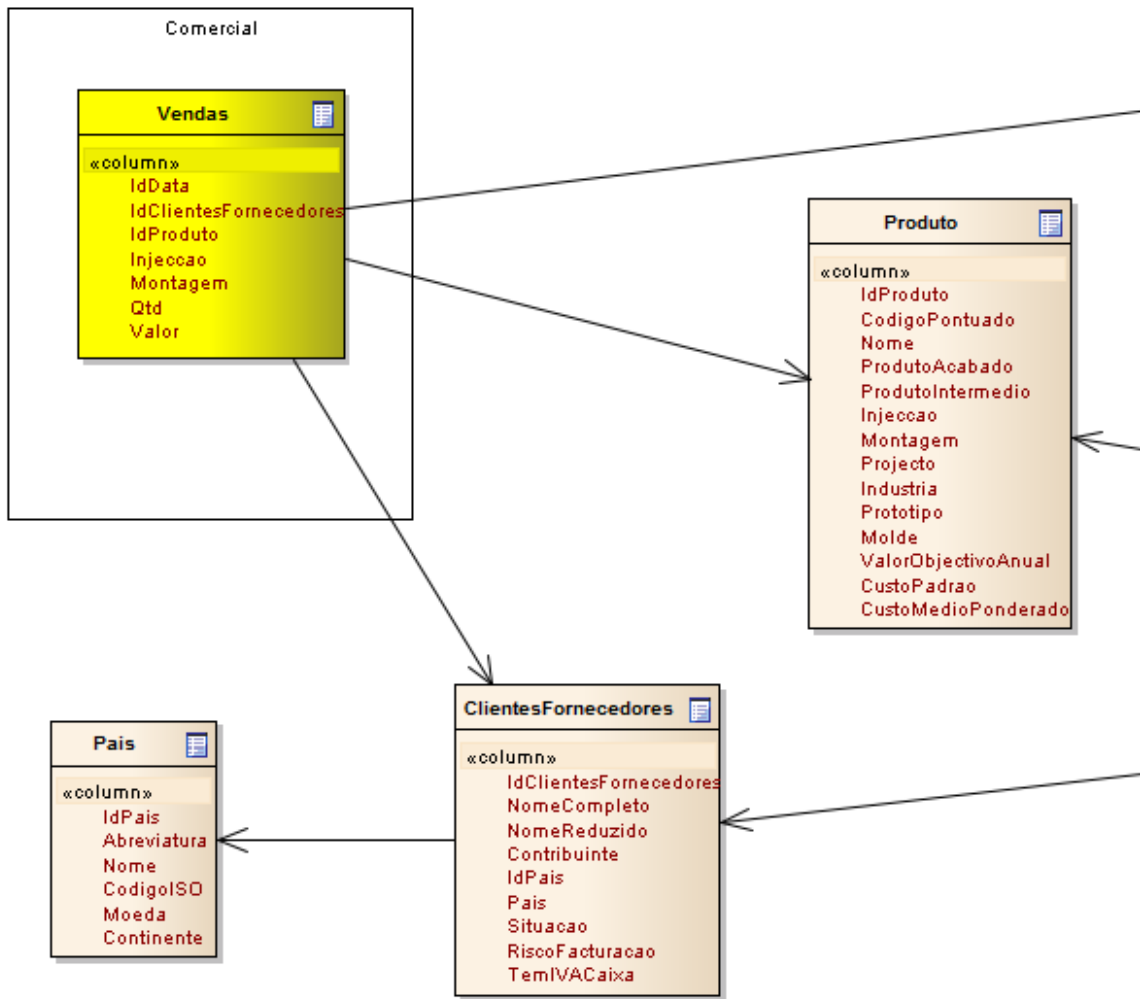


Figura 6.5. Detalhe do modelo dimensional para a área Comercial

A figura 6.5 mostra o detalhe da área Comercial, com a tabela de factos “Vendas” e com as dimensões “ClientesFornecedores”, “Produto”, “Pais” e “Data”. Pretende-se a breve prazo estender a data warehouse a outras áreas da empresa, pelo que a opção de “estender” a tabela “Pais” para fora da “ClientesFornecedores” (processo chamado de snowflake) foi consciente, pois há necessidade de partilhar esta tabela. Este detalhe mostra como a data warehouse responde a todas as questões que o departamento Comercial pretende ver respondidas.

Na figura 6.6, podemos observar um detalhe do modelo dimensional para a área da Logística.

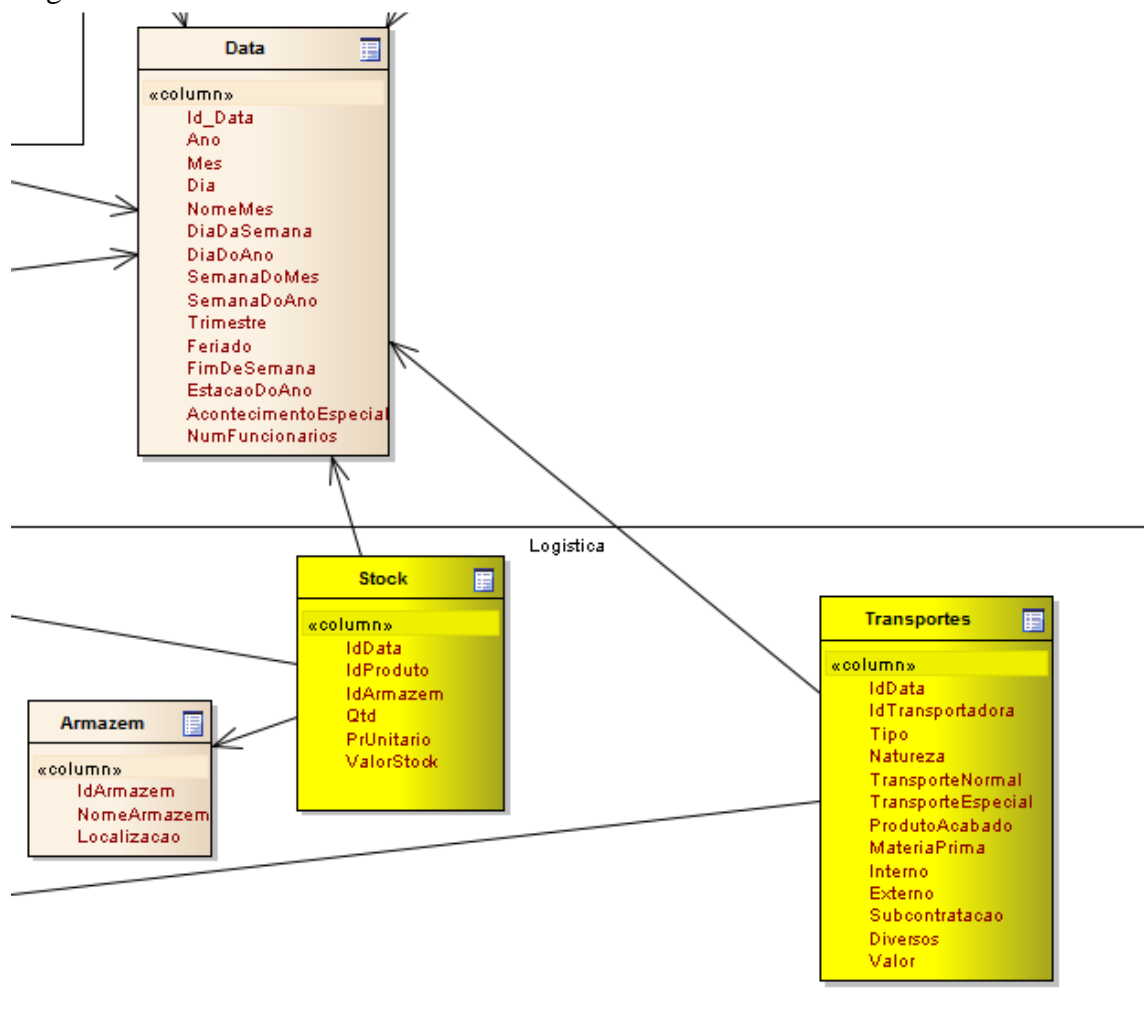


Figura 6.6. Detalhe do modelo dimensional para a área da Logística

A figura 6.6 mostra o detalhe da área da Logística, com as tabelas de factos “Stock” e “Transportes”, e com as dimensões “Armazem” e “Data”, que defina a sua granularidade. Desta forma respondemos a todas as questões que o departamento de Logística pretende ver respondidas.

Neste capítulo apresentámos a empresa e o seu negócio, analisámos as questões a que era necessário dar resposta, e definimos o modelo dimensional para a data warehouse.

No capítulo seguinte, abordaremos a implementação. Vamos instalar as ferramentas necessárias, realizar o ETL, e configurar a SpagoBI para dar resposta às necessidades de extracção de informação.

7. IMPLEMENTAÇÃO

Neste capítulo abordaremos a implementação da suite open source SpagoBI em ambiente industrial real. Tendo a suite escolhida e definido o modelo dimensional, vamos descrever neste capítulo a implementação.

Na MICROplásticos privilegia-se o software open source, pelo que vários serviços estão assentes em sistemas operativos Linux.

Tendo em conta este facto, optou-se por preparar uma máquina Linux virtualizada que facilmente se colocou a funcionar no ambiente real, integrada na restante infra-estrutura da empresa. Optou-se pelo OpenSuse (OpenSuse, 2014) por ser também este já usado e conhecido na empresa.

Há muitas distribuições já com pré-instalação de diversas ferramentas disponíveis para download na Internet. No caso, optou-se por fazer download duma imagem virtualizada, preparada para o VMWare (VMWare, 2014), no site do Traffic Tool (TrafficTool, 2014), pronta para ser executada através do VMWare Player (Player, 2014).

7.1. Instalação da SpagoBI

A versão mais recente da suite SpagoBI é a 4.1.

A sua instalação em Linux pode ser consultada em anexo.

Feita a instalação, estamos prontos a explorar a ferramenta, abrindo o browser através do endereço <http://localhost:8080/SpagoBI>, e fazendo login, como mostra a figura 7.1.



Figura 7.1. SpagoBI no browser, preparado para fazer login

Na secção seguinte abordaremos a criação da data warehouse sobre o motor MySQL, sendo o motor de base de dados escolhido para a suportar,

7.2. MySQL

Para carregar os dados para a data warehouse o motor de base de dados escolhido foi o MySQL, por ser suportado pela SpagoBI, por ser gratuito e um dos mais utilizados em todo o mundo, e também porque é um dos motores de base de dados usados na empresa. Seguindo o modelo de dados definido, usou-se o MySQL Workbench para desenhar o modelo entidade-relacionamento, como pode ser observado na figura 7.2.

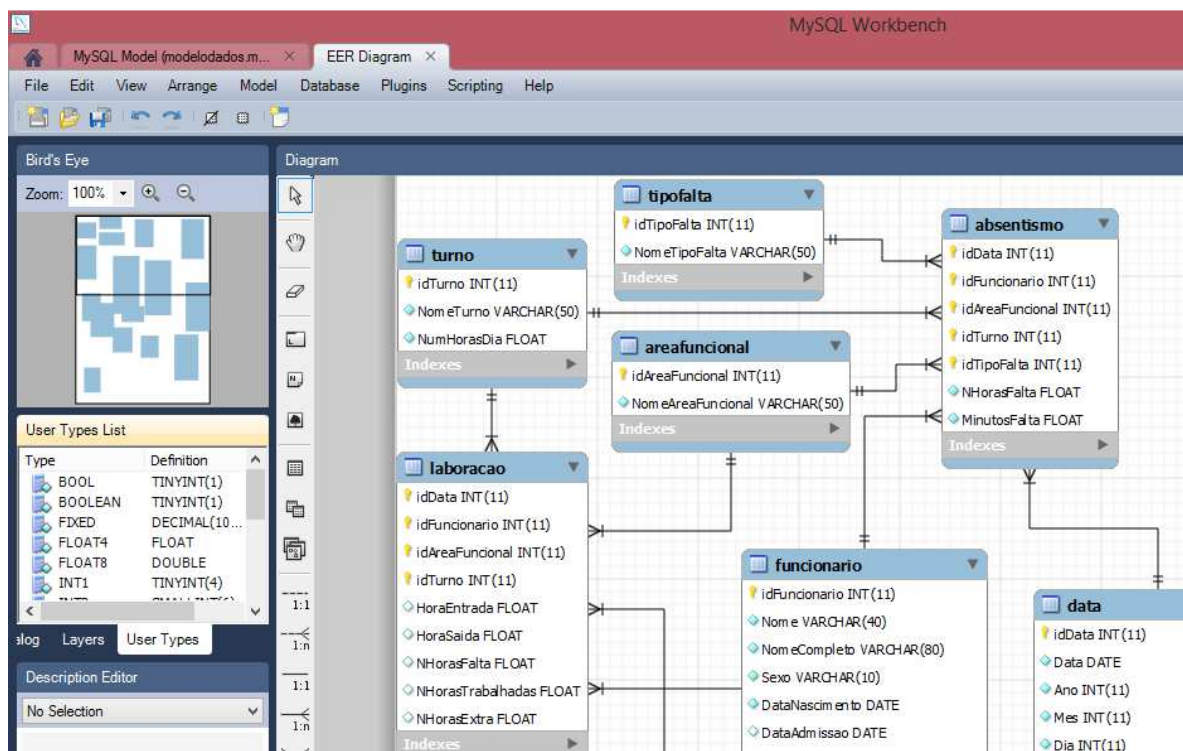


Figura 7.2. Modelo entidade-relacionamento no MySQL Workbench

Usando a função “Forward Engineer” no menu “Database”, e seguindo o *wizard*, foi gerada a base de dados pretendida, como mostra a figura 7.3 para o caso específico da tabela “laboracao”.

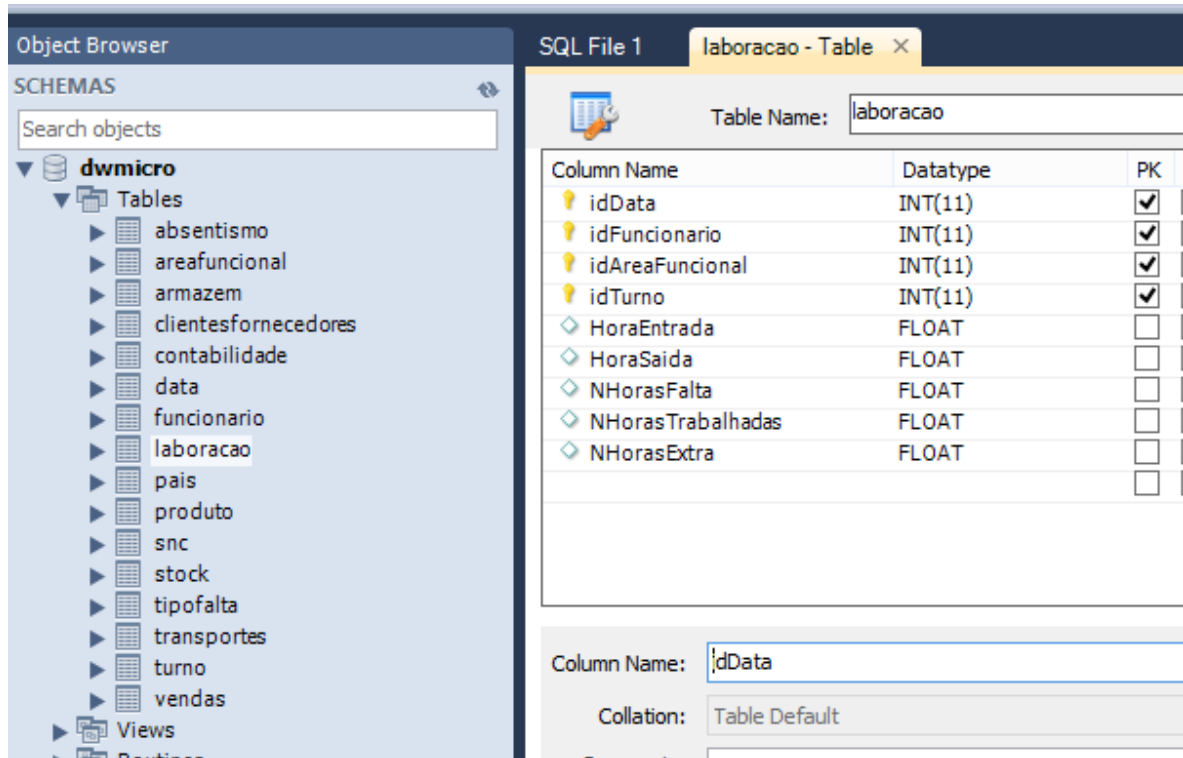


Figura 7.3. Data warehouse gerada através do MySQL Workbench

Tendo a data warehouse criada, temos tudo preparado para partir para a fase seguinte, e uma das mais importantes de todo o processo de BI, o ETL. Na secção seguinte abordaremos como foi realizada esta fase primordial na implementação.

7.3. Talend

A SpagoBI suporta automatização de ETL a partir da ferramenta Talend (Talend, 2014), que instalámos e usámos para para realizar o mesmo.

Esta ferramenta é suportada por várias suites, e pela experiência que tivemos pode-se afirmar que é simples e intuitiva, e com um óptimo suporte de documentação online (TalendDocs, 2014).

A base de dados operacional na empresa assenta sobre um motor Oracle. Assim e visto que para os vários *jobs* necessitaríamos de introduzir componentes de consulta à base de dados operacional, e de consulta e escrita à data warehouse, optámos por criar as ligações às bases de dados no repositório da ferramenta, no caso em Metadata (Metadata, 2014). As queries de consulta também foram aí guardadas, sendo mais fácil de gerir e propagar quaisquer alterações para os diversos *jobs* onde estejam a ser usadas, como podemos observar na figura 7.4.

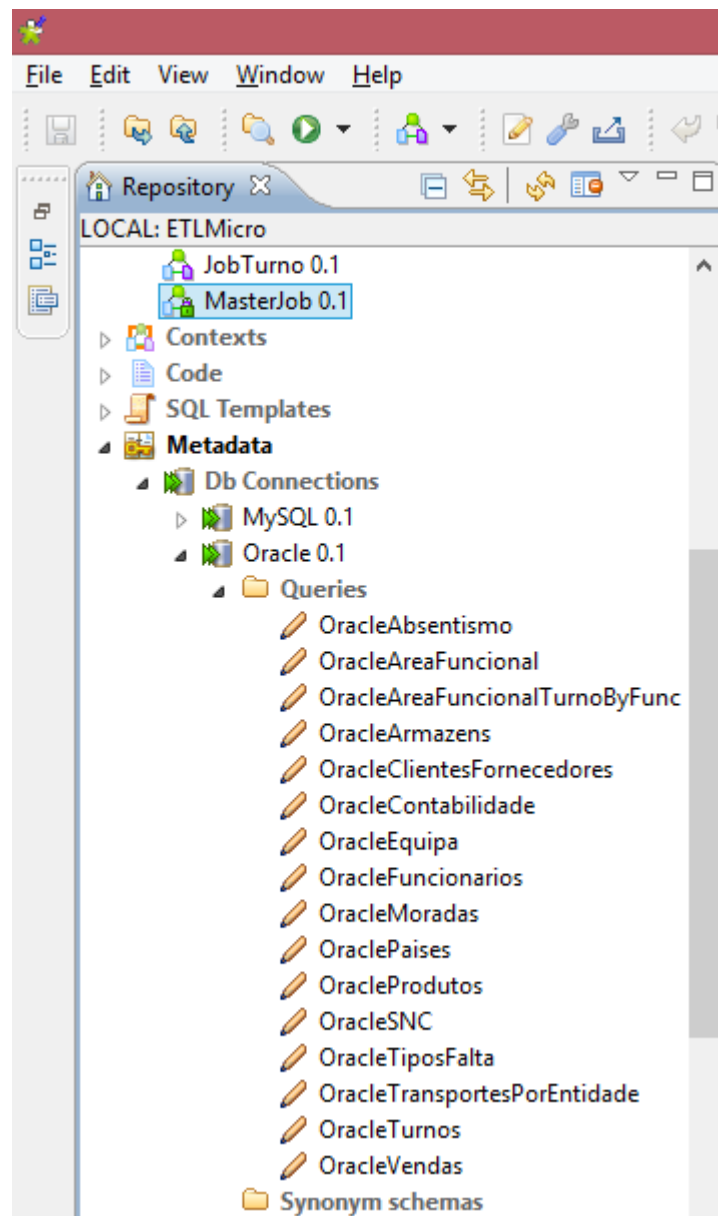


Figura 7.4. Ligações e queries no repositório do Talend

Na figura 7.5 abaixo pode observar-se a configuração da ligação à base de dados operacional, que ficará disponível para usar nos *jobs* depois de pressionar a opção “Export as context”.

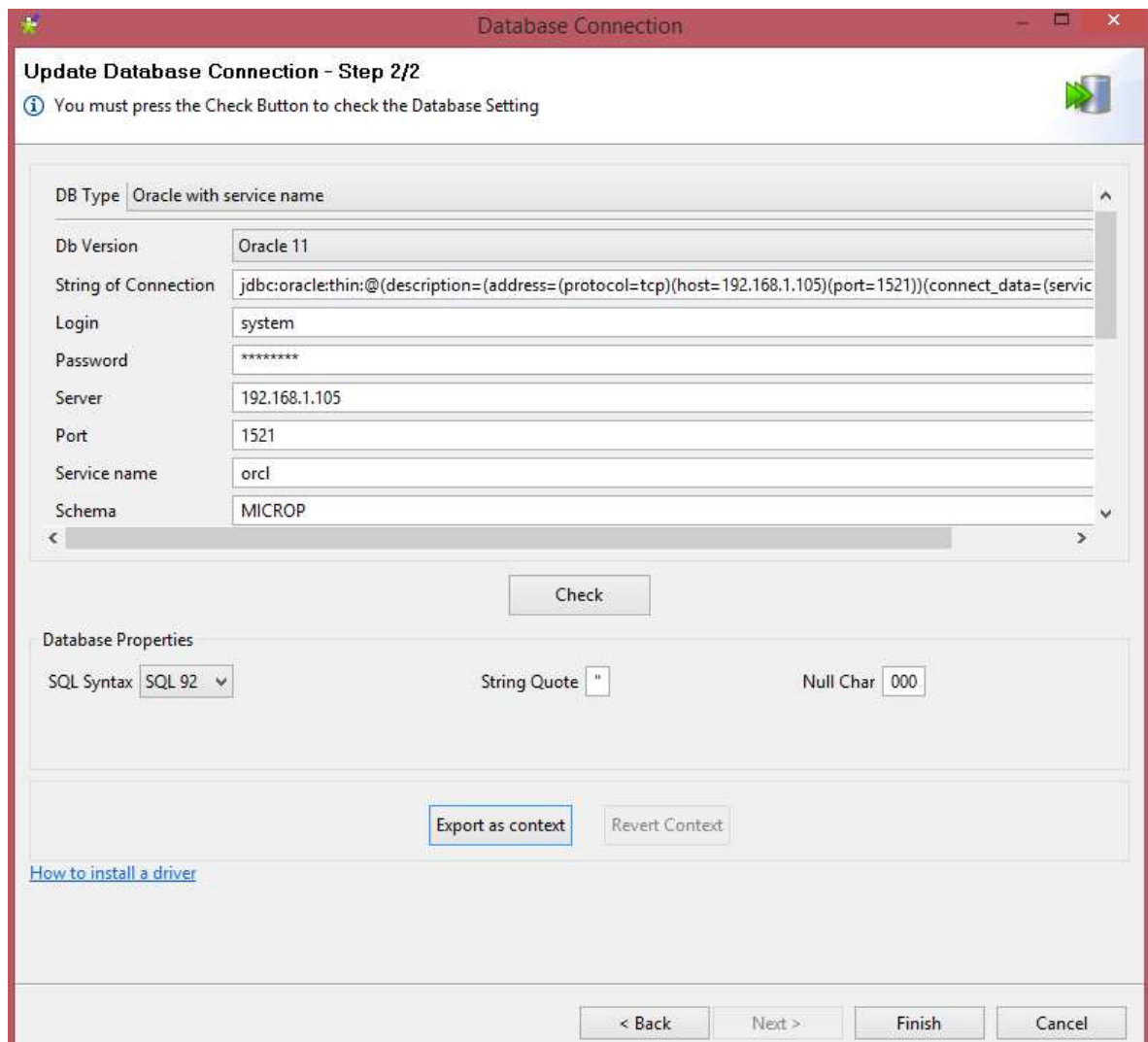


Figura 7.5. Configuração da ligação à base de dados operacional, Oracle

Para uma fácil manutenção foi criado um *job* para cada tabela a preencher na data warehouse. Mas para os agregar e executar, criámos um *job* principal que tem essa responsabilidade para com os *subjobs*, como pode ser observado na figura 7.6.

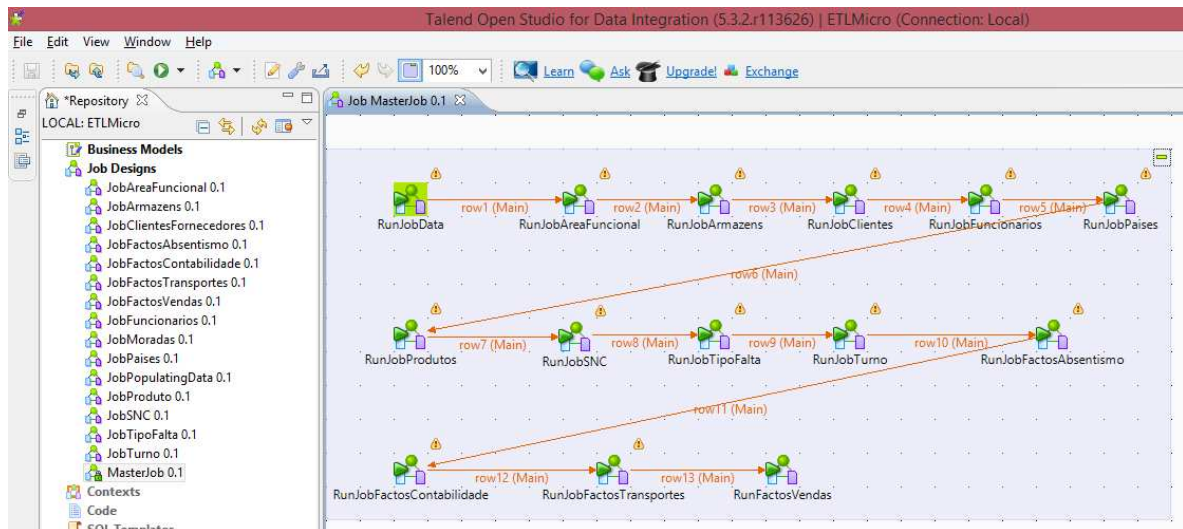


Figura 7.6. Job principal no Talend, contendo a sequência de execução dos *sub-jobs*

Vamos explicar o processo de ETL para a tabela de factos Absentismo, que podemos observar na figura 7.7.

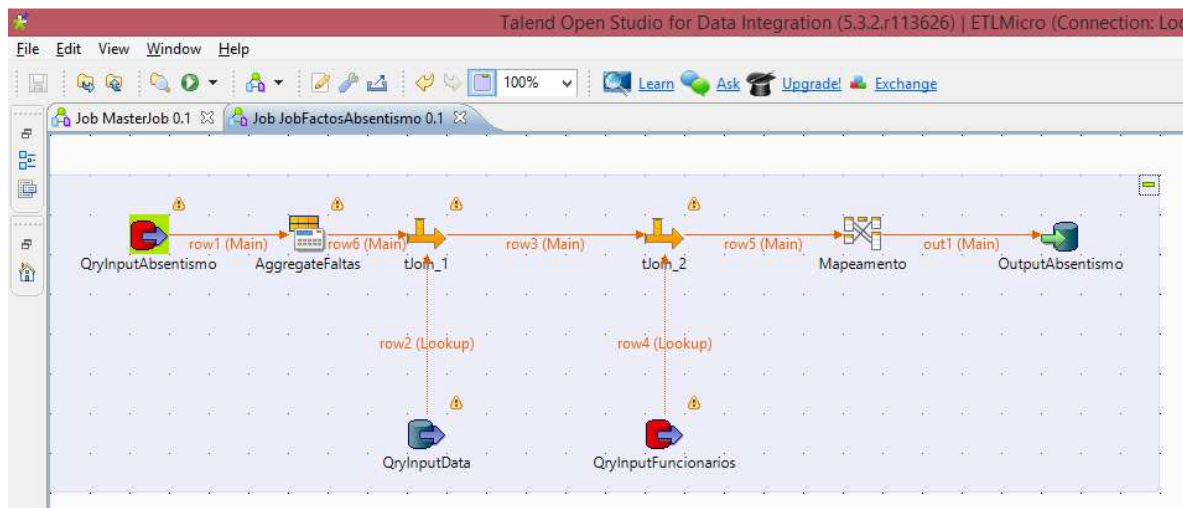


Figura 7.7. Processo de ETL para a tabela de factos Absentismo

Explicando este processo, os dados são lidos através do componente “QryInputAbsentismo”, que usa a ligação à base de dados Oracle, e a query de consulta que guardámos no repositório, como pode ser observado na figura 7.8.

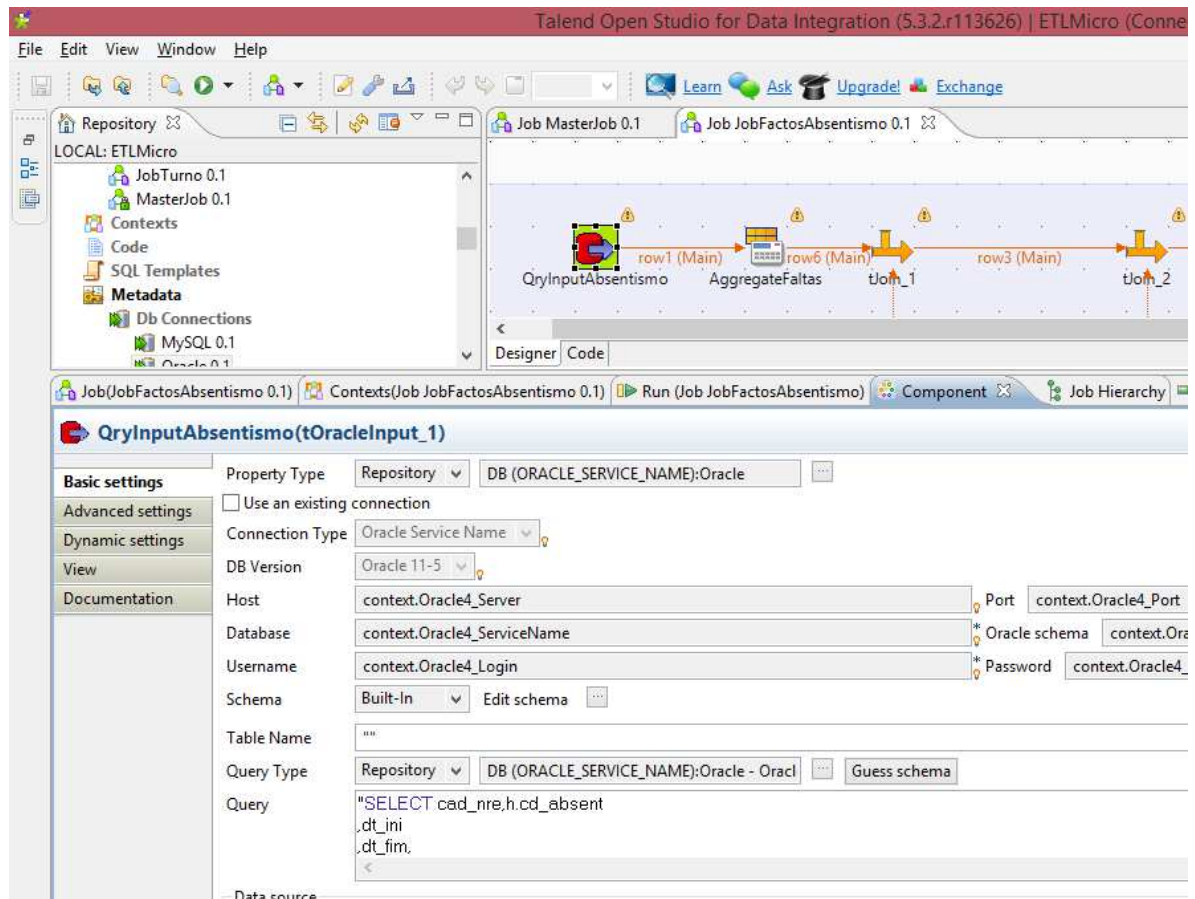


Figura 7.8. Configuração do componente QryInputAbsentismo

Este componente liga ao “AggregateFaltas”, que soma as horas e minutos de falta segundo os critérios de agregação definidos, que são a Data, o Tipo de Falta, e o funcionário, como pode ser observado na figura 7.9. Esta agregação é necessária por existirem casos de mais que um registo para o mesmo funcionário para as mesmas condições de agregação.

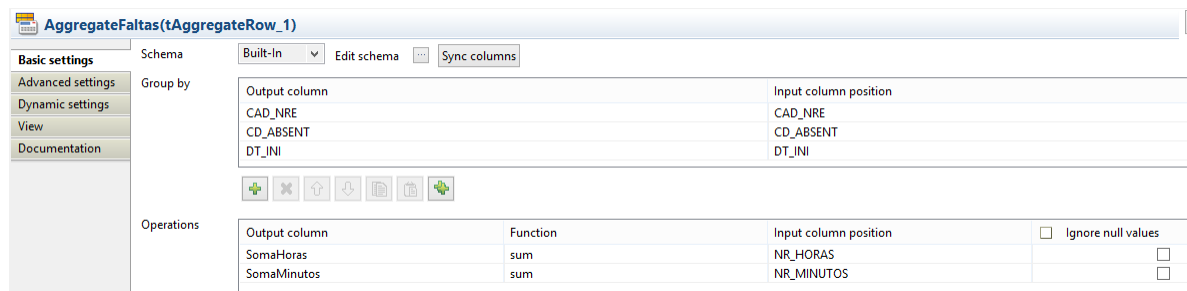


Figura 7.9. Configuração do componente AggregateFaltas

O componente seguinte faz um *lookup* à tabela Data da data warehouse e devolve o *ID* que atribuímos à data da falta, como podemos observar na figura 7.10.

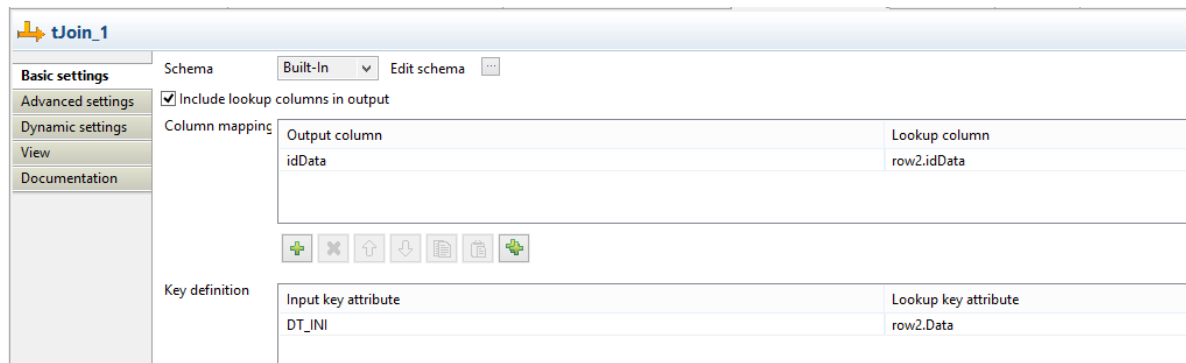


Figura 7.10. Configuração do componente de lookup à tabela Data

No componente seguinte fazemos novo *lookup*, no caso à base de dados operacional para obter a Área Funcional e o Turno que o funcionário estava no momento da falta, como pode ser observado na figura 7.11.

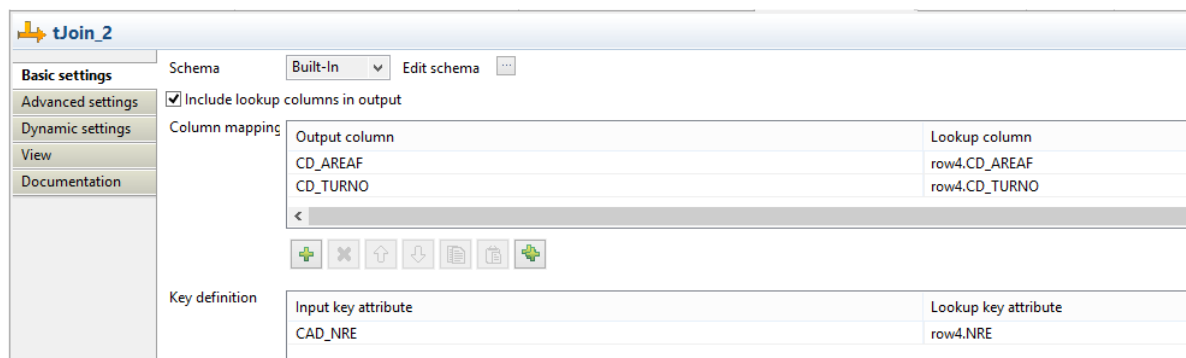


Figura 7.11. Configuração do componente de lookup para os dados do funcionário

O passo seguinte foi realizar o mapeamento para poder inserir os dados na data warehouse. A figura 7.12 mostra as relações realizadas no componente “Mapeamento”, bem como as expressões de transformação de dados usadas. A “row5” serve de entrada no mapeamento, e tem preenchidos os campos que vêm de trás. Assim, mapeámos e transformámos os dados da seguinte forma para o *output* que preencherá a tabela “absentismo”:

- row5.idData – mapeado com idData
- row5.CAD_NRE – mapeado com idFuncionario
- row5.CD_AREAF – transformado em inteiro, mapeado com idAreaFuncional
- row5.CD_TURNO – mapeado com idTurno
- row5.CD_ABSENT – transformado em inteiro, mapeado com idTipoFalta
- row5.SomaHoras – mapeado com NHorasFalta
- row5.SomaMinutos – mapeado com MinutosFalta

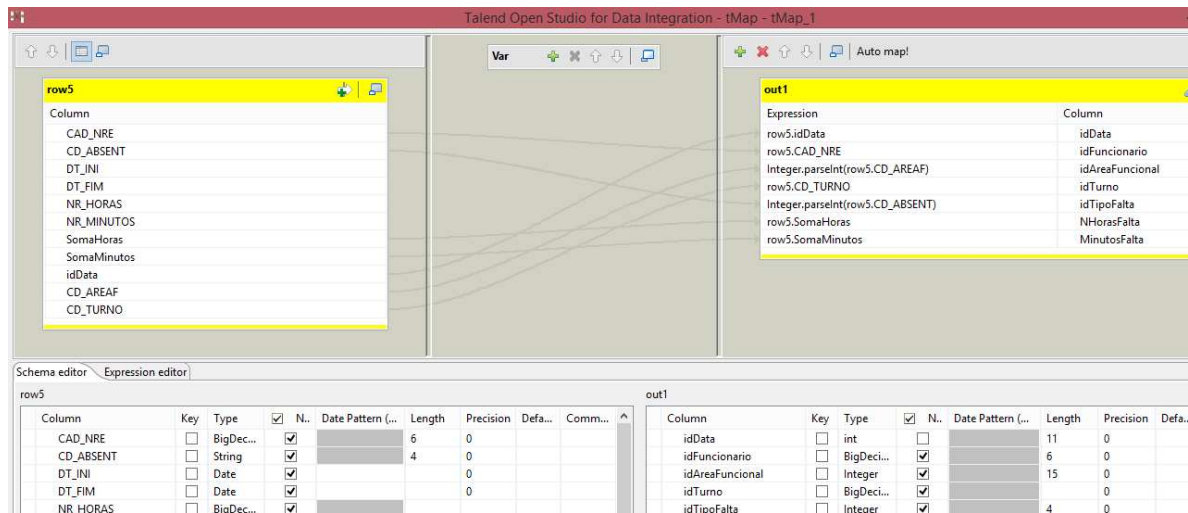


Figura 7.12. Configuração mapeamento para futura gravação de dados

Por fim foi usado um componente de *output* para o MySQL, usando a configuração do repositório, e que, no caso, apenas indicámos a tabela a preencher, visto toda a parametrização e mapeamento já estar realizado, e com a configuração que pode ser observada na figura 7.13.

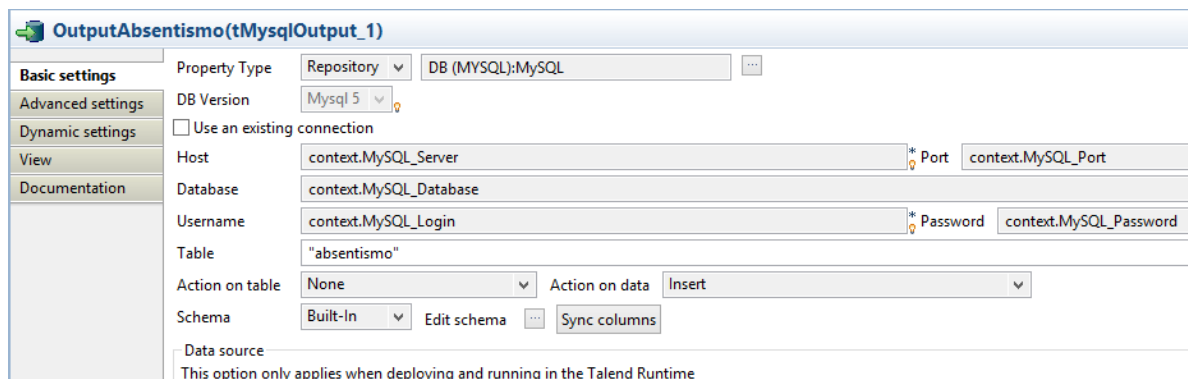


Figura 7.13. Configuração do componente responsável por gravar os dados

No MySQL Workbench podemos consultar os dados inseridos por este processo de ETL, como pode ser observado na figura 7.14.

The screenshot shows the MySQL Workbench interface. At the top, there are tabs for 'SQL File 1' and 'Query 2'. Below the tabs is a toolbar with various icons. The main area displays a SQL query: `SELECT * FROM dwmicro.absentismo;`. Below the query editor is a 'Filter:' field and a toolbar with icons for 'Edit', 'File', 'Autosize', and 'Fetch rows'. The bottom part of the screenshot shows a table with the following data:

	idData	idFuncionario	idAreaFuncional	idTumo	idTipoFalta	NHorasFalta	MinutosFalta
	1098	964	117	6	108	3	180
	1099	182	117	15	106	7.5	450
	1099	739	117	13	101	7.5	450
	1100	182	117	15	106	7.5	450
	1100	458	120	0	108	2.5	150
	1100	641	117	14	107	1	60
	1100	668	117	4	108	1.25	75
	1100	696	117	14	108	0.5	30
	1100	698	117	14	108	1	60
	1100	853	117	2	108	1	60
	1100	864	117	4	108	1	60
	1100	871	117	4	108	1	60
	1100	901	117	15	108	7.5	450
	1100	906	117	2	108	1	60
	1100	948	117	0	108	1	60

Figura 7.14. Visualização dos dados inseridos no MySQL Workbench

O mesmo tipo de método foi seguido para as restantes tabelas de factos e dimensões. Por necessidade de alteração do processo interno da empresa, visto não haver um registo fiável em qualquer suporte digital, foi suspenso de momento o ETL e conseqüente obtenção de indicadores relativamente ao ponto de pessoal. Relativamente aos restantes dados para indicadores, todos foram carregados para a data warehouse como estava previsto no modelo de dados.

7.4. Extração da informação

Já foi referido que o suporte à criação de documentos de extração de informação é realizado pelo SpagoBI Studio. Depois de instalado, recorreremos a esta ferramenta para desenhar os documentos analíticos que necessitamos.

A documentação sobre esta ferramenta pode ser encontrada na wiki da SpagoBI (WikiStudio, 2014), ou através de Ogutu que explica minuciosamente a criação de muitos destes documentos.

Na figura 7.15 podemos observar a configuração de um documento OLAP, em que identificamos como “cubo” a tabela de factos “Absentismo”, e como dimensões as tabelas “Funcionario” e “Data”.

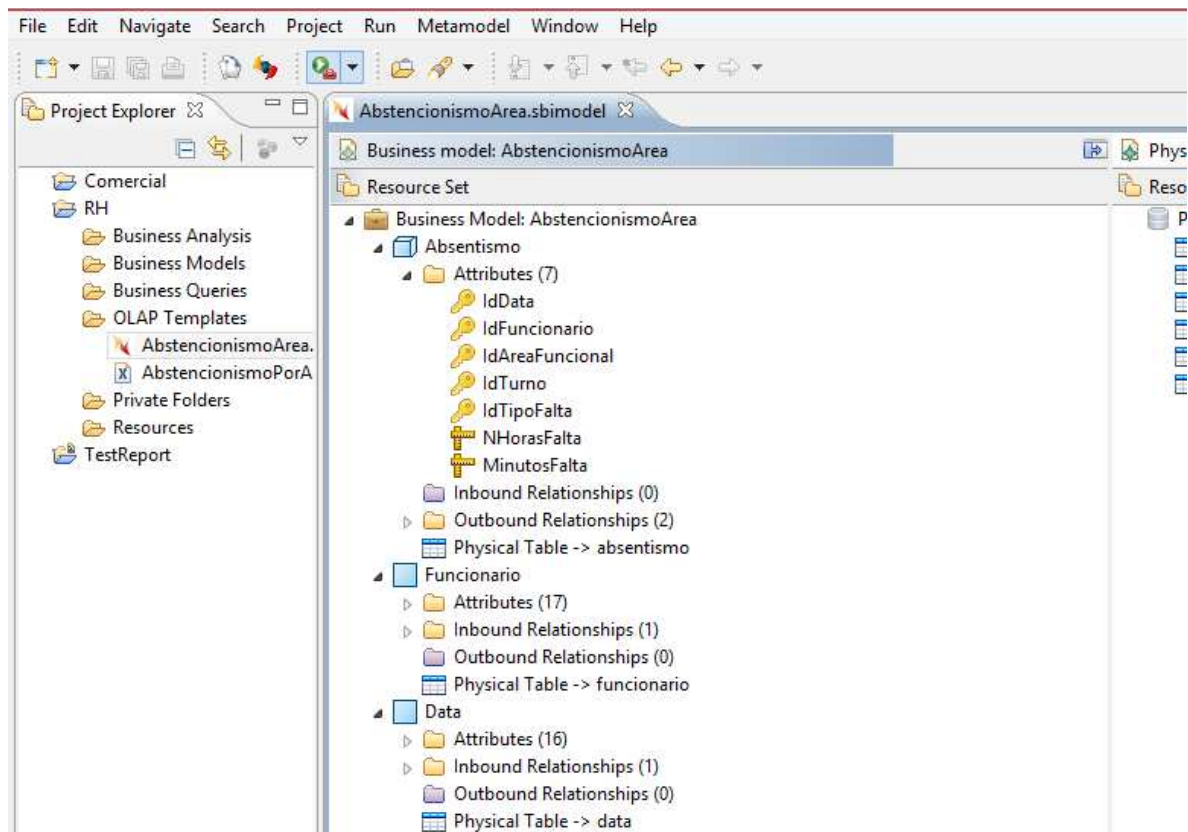


Figura 7.15. Configuração de um documento OLAP no SpagoBI Studio

No “cubo” Absentismo necessitamos identificar os factos mensuráveis e a sua relação com as tabelas de dimensão. Nas dimensões é necessário indicar a hierarquia para o *drill down* e quais os campos que compõem essa hierarquia como pode ser observado na figura 7.16.

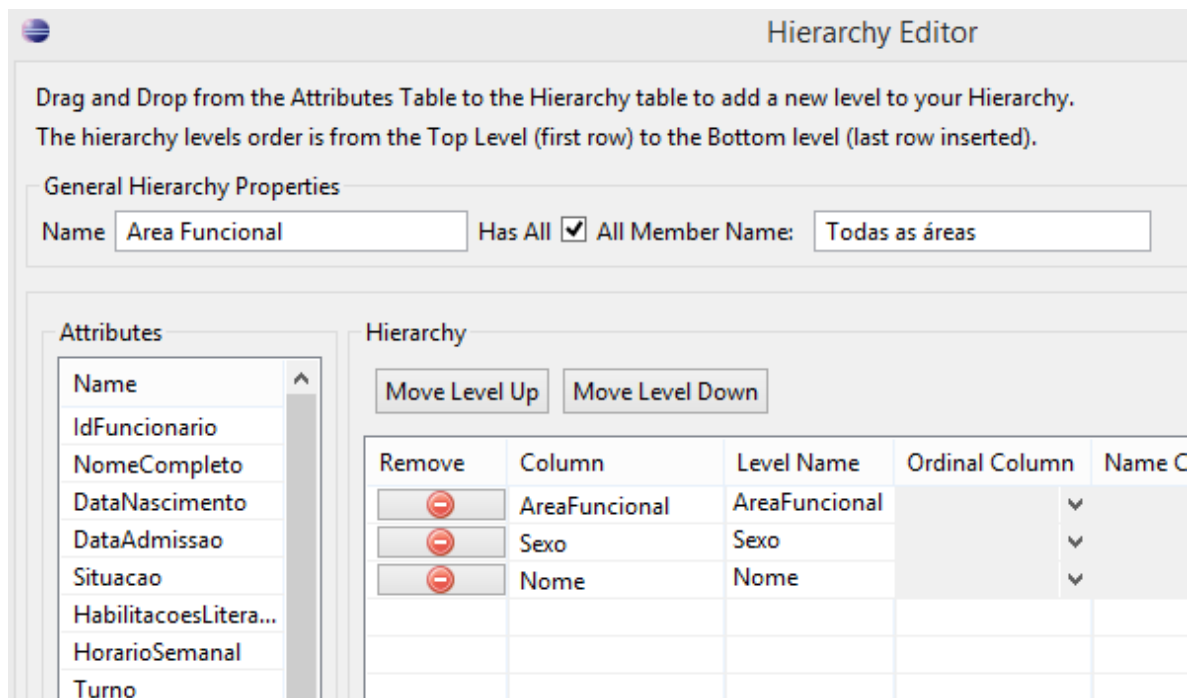


Figura 7.16. Hierarquia da pesquisa OLAP na Area Funcional

Depois de efectuadas as definições, é necessário gerar o ficheiro *xml* que servirá de template para a exploração da informação, e fazer o upload para o servidor da SpagoBI, como mostra a figura 7.17.

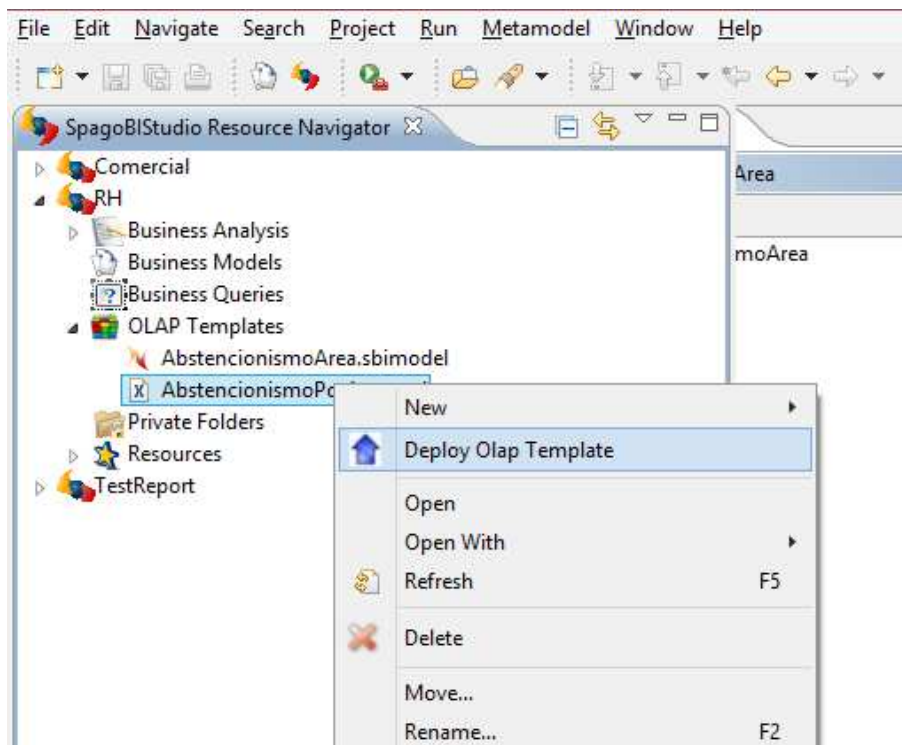


Figura 7.17. Upload do template OLAP para o servidor SpagoBI

No SpagoBI podemos então seleccionar o documento que gerámos e explorar a informação tal como foi definida, como pode ser visualizado na figura 7.18.

Funcionario	Data	Measures NHorasFalta
-Todas as áreas	+2013	16,430
+?	+2013	168
+Administração	+2013	
+Arm. Matéria Prima	+2013	8
+Arm. Prod Acabado	+2013	23
+CAD	+2013	
+Controlo Qualidade	+2013	1,310
+Dept Admin/Financ	+2013	
-Dept Comercial	+2013	218
-Feminino	+2013	218
Anabela Jordão	+2013	
Dora Batata	+2013	209
Marisa Oliveira	+2013	
Silvia Teles	-2013	9
	+1	
	+2	8
	+3	
	+4	
	+5	
	+6	1
	+7	
	+8	
	+9	
	+10	
	+11	
	+12	
+Dept Engenharia	+2013	324
+Dept Fabril	+2013	

Figura 7.18. Explorar o documento AbsentismoArea no SpagoBI

Para os diferentes indicadores e informação a explorar nas diversas áreas em que intervimos foram criados os templates com suporte ao SpagoBI Studio, e criados os diversos documentos.

Por exemplo, para criar um documento de geolocalização com recurso a um mapa, é necessário ter um ficheiro com o mapa pretendido no formato *.svg (Scalable Vector Graphics). No caso da implementação com o mapa do mundo fizemos download do site

Wikimedia Commons (Wikimedia, 2014), em que cada forma geométrica vectorial tem um identificador que deve estar relacionado com a nossa origem de dados, como pode ser observado na figura 7.19.

```
<g
  id="Countries">
  <path
    id="portugal"
    d="M430.93,211.241-0.62,8.651-1.77,1.610.18,0.9811.24,2.051-0.8,2.511.33,0.4513.1-0.3
  <path
    id="espanha"
    d=
    "M448.36,205h-12.741-2.57-1.161-1.24,0.091-1.5,3.1210.53,3.2114.87,0.4510.62,2.051-2.
    .9916.9-6.2410.35-2.761-6.28-0.09L448.36,205L448.36,205z"/>
```

Figura 7.19. Formato do ficheiro SVG do mapa do mundo

Depois de carregado o mapa para a plataforma e de gerado o template no SpagoBI Studio, no qual indicamos os campos de identificação dos países, e qual o campo de valor na nossa origem de dados, criamos o documento na plataforma, como pode ser observado na figura 7.20.

The screenshot shows the SpagoBI web interface at the URL `localhost:8080/SpagoBI/servlet/AdapterHTTP?PAGE=LoginPag`. The main content area is titled 'DOCUMENT DETAILS' and contains a form with the following fields and values:

Label	Vendas por Pais 2013 *
Name	Vendas por Pais 2013 *
Description	
Type	Location Intelligence
Engine	Geo Engine
Data Source	DWMicro
Dataset	VendasPorPais
State	Development
Community	
Scope	Private
Refresh seconds	0
Criptable	<input type="radio"/> True <input checked="" type="radio"/> False
Visible	<input checked="" type="radio"/> True <input type="radio"/> False
Visibility restrictions	
Preview file	Escolher ficheiro Nenhum ficheiro selecionado
Template	Escolher ficheiro Mapa2.xml

The SpagoBI logo is visible in the bottom left corner of the interface.

Figura 7.20. Criação do documento “Vendas por Pais 2013” no SpagoBI

O resultado final deste documento pode ser observado na figura 7.21 abaixo.

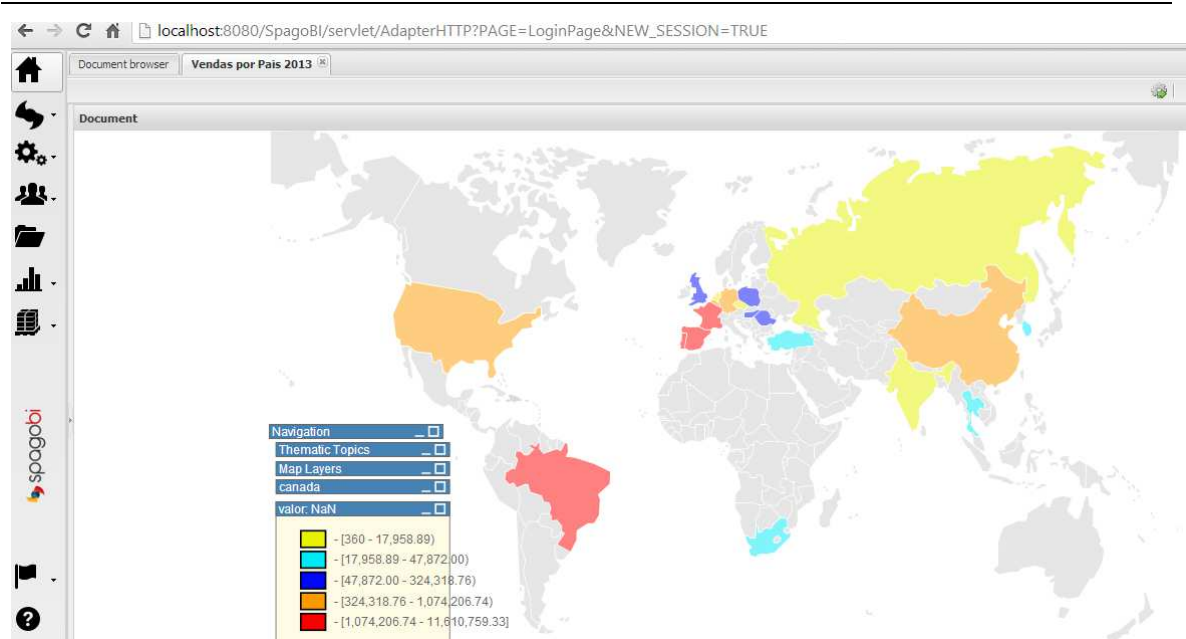


Figura 7.21. Resultado final do documento “Vendas por Pais 2013”

Outro exemplo de extração de informação gerada são gráficos, de que a figura 7.22 abaixo é representativa para a área da Administração. Neste caso são mostrados os pesos dos gastos com pessoal relativamente às vendas.

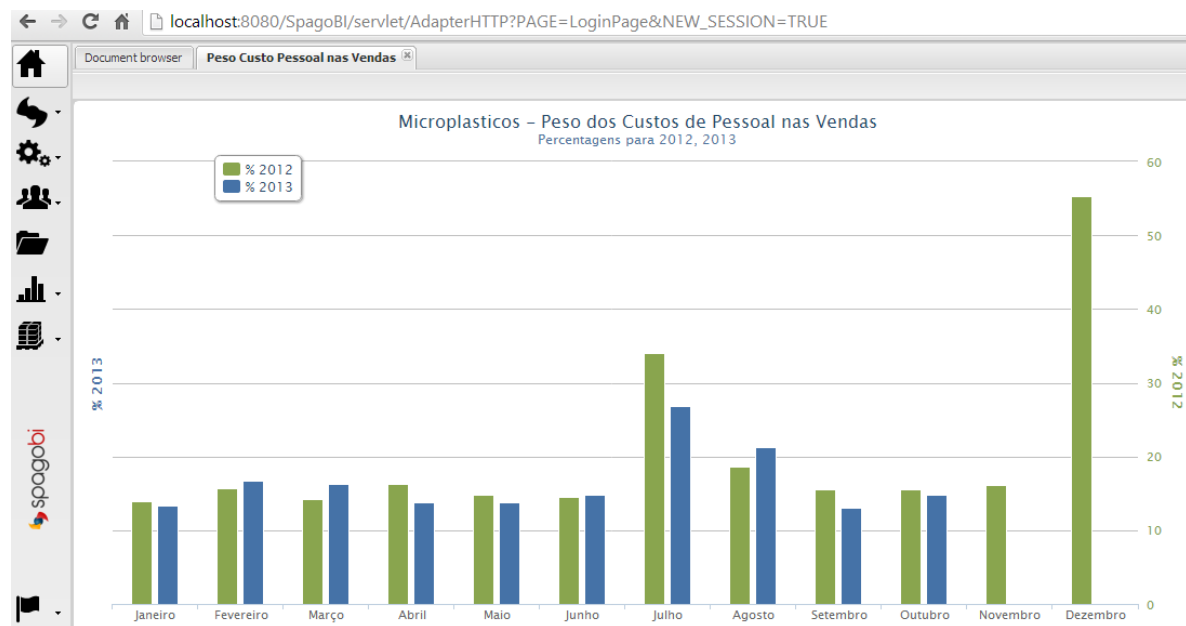


Figura 7.22. Gráfico do peso do Custo do Pessoal relativamente às Vendas

Para todos os indicadores pretendidos foram elaborados os mais diversos documentos e mapas por forma a responder às várias necessidades da empresa, que auxiliam o seu acompanhamento, contribuindo para o conhecimento próprio e tornando este numa mais valia competitiva. Tomando como exemplo a figura 7.22 acima, facilmente se percebe

que os custos com pessoal relativamente às vendas variam entre os 12 e os 16%, com exceção dos meses em que há lugar a pagamento de subsídios, e do mês de Agosto em que a empresa fecha temporariamente para férias e os custos com pessoal têm um peso maior. Olhando para o gráfico e numa análise muito superficial, percebe-se que não há desequilíbrios, o que indica que é uma empresa estável, e de “boa saúde”.

7.5. Acessos

As configurações de regras, autenticação e acessos aos diversos documentos foram realizadas para os diversos utilizadores que têm acesso a este tipo de informação, através do menu de gestão que pode ser observado na figura 7.23.

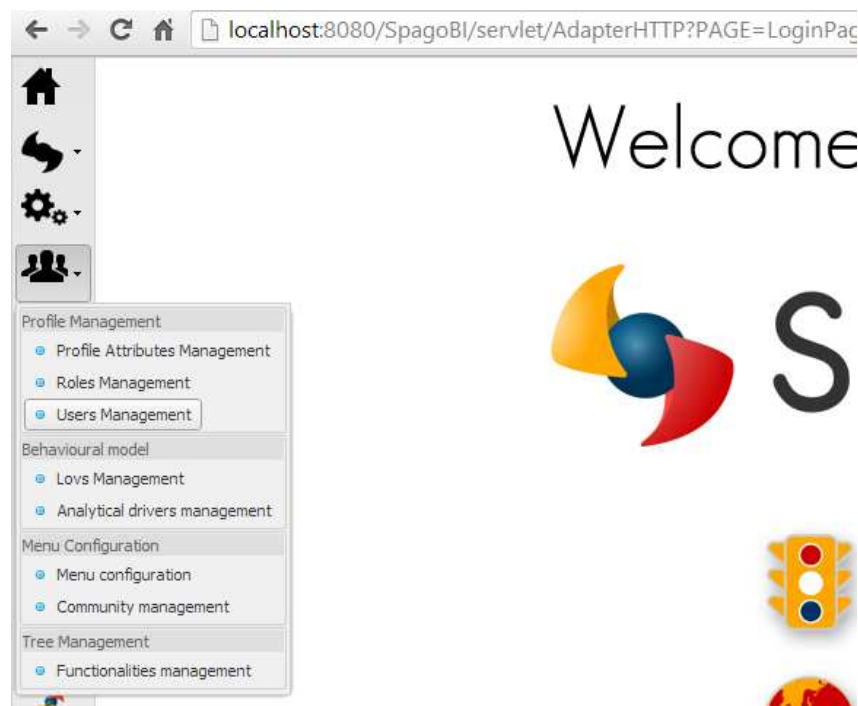
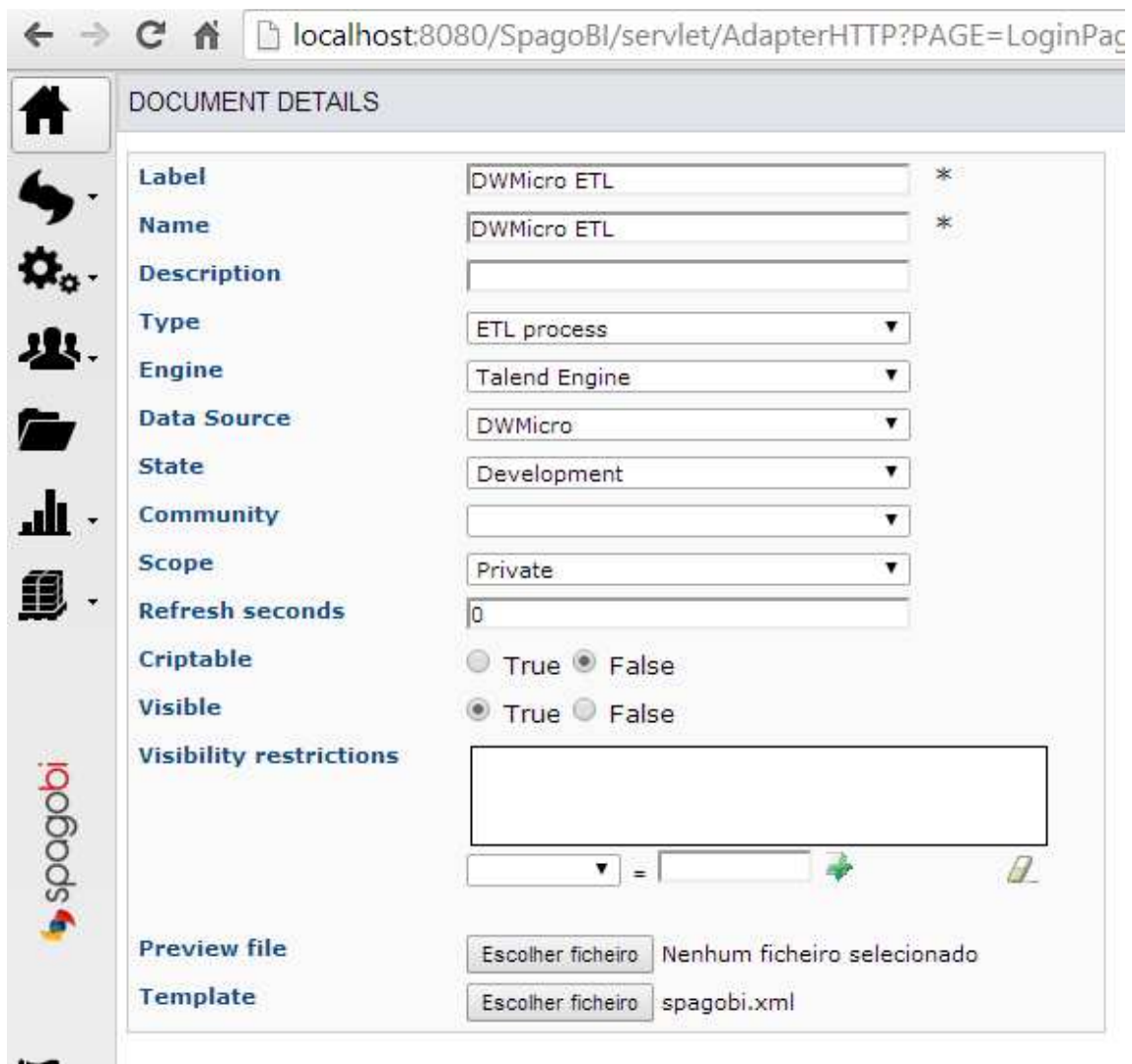


Figura 7.23. Menu da gestão de utilizadores no SpagoBI

Esta é uma parte fundamental numa implementação deste tipo, pois estamos a lidar com informação normalmente confidencial para a maioria dos colaboradores, e em que cada utilizador do sistema terá acesso restrito a determinados documentos. Foram criados grupos de utilizadores, com restrições de visualização de pastas por grupo, e a estes grupos foram associados os utilizadores que têm acesso à ferramenta. A restrição foi feita ao grupo de utilizador tornando a manutenção mais fácil.

7.6. Automatização

Para automatizar o carregamento do ETL foi também criado um documento do tipo “ETL Process”, com recurso ao “Talend Engine”, como mostra a figura 7.24, e que foi agendado para uma execução diária.



The screenshot displays the SpagoBI web interface for creating an ETL document. The browser address bar shows the URL: localhost:8080/SpagoBI/servlet/AdapterHTTP?PAGE=LoginPag. The page title is 'DOCUMENT DETAILS'. The form contains the following fields and options:

Label	DWMicro ETL *
Name	DWMicro ETL *
Description	
Type	ETL process
Engine	Talend Engine
Data Source	DWMicro
State	Development
Community	
Scope	Private
Refresh seconds	0
Criptable	<input type="radio"/> True <input checked="" type="radio"/> False
Visible	<input checked="" type="radio"/> True <input type="radio"/> False
Visibility restrictions	
Preview file	Escolher ficheiro Nenhum ficheiro selecionado
Template	Escolher ficheiro spagobi.xml

Figura 7.24. Criação do documento de carregamento do ETL no SpagoBI

Desta forma fechámos o ciclo da implementação pretendida, que no nosso entender responde às necessidades levantadas e confere à empresa um conjunto de ferramentas de análise que lhe traz uma vantagem competitiva que advém do facto de facilitar o seu próprio conhecimento e do seu negócio.

Esta solução ainda está neste momento numa fase de avaliação e validação por parte dos responsáveis dos vários departamentos onde houve intervenção na empresa, existindo ainda a necessidade de alterar alguns processos internos para a implementação entrar em pleno funcionamento.

8. CONCLUSÕES E TRABALHO FUTURO

O problema principal que nos propusemos resolver com este trabalho foi levar o Business Intelligence para as empresas sem custos. Assim, este trabalho de investigação teve como objectivo seleccionar e implementar uma suite open source de Business Intelligence numa empresa de média dimensão, tentando demonstrar que estas soluções são viáveis neste tipo de empresas, sem que tal implique um significativo aumento de custos na sua aquisição e implementação.

Vivendo-se actualmente um ambiente de alguma instabilidade no mercado empresarial um pouco por todo o mundo, todo o conhecimento que as empresas adquiram sobre o seu negócio pode traduzir-se numa vantagem competitiva face à concorrência. Nestes momentos de instabilidade impõe-se mais conhecimento e menor margem de erro nas decisões estratégicas.

Disponibilizar atempadamente o maior número de informações e conhecimento, está na base da escolha das alternativas que se levantam como suporte à decisão, contribuindo assim para um incremento do valor do negócio.

O Business Intelligence, como mostrado ao longo deste trabalho, é a tecnologia que suporta esta decisão disponibilizando o conhecimento necessário. O BI deve ser bem estruturado e pensado por parte dos implementadores e analistas, num trabalho que deve ser desenvolvido com os gestores e utilizadores finais. Estes últimos terão então disponível um conjunto de elementos que pode apoiar na tomada de decisão e acompanhamento do seu negócio, sendo certo, que o BI por si só não resolve quaisquer problemas.

O facto de se implementarem soluções open source que garantam um determinado conjunto de garantias de satisfação, ajuda a que as empresas possam optar por estas soluções, pois traduz-se num benefício entre as vantagens que oferece e os custos que implica.

Tendo em conta estas questões, e após analisar o estado da arte relativamente ao open source, ao Business Intelligence e às suites open source BI existentes no mercado, foram identificadas aquelas que poderiam ser uma solução para o problema colocado.

Das cinco seleccionadas (Jaspersoft, Palo, Pentaho, Spago BI e Vanilla), foram identificadas a Pentaho e a Spago BI como sendo as melhores soluções no contexto da análise.

Com vista a eleger entre as soluções seleccionadas aquela que iríamos implementar, foi apresentado um método de avaliação de software open source, o Open BRR, que serviu para realizar a avaliação das suites, e seleccionar a melhor para a implementação. Após avaliação diversos critérios na implementação da metodologia, resultou na escolha da SpagoBI como sendo a melhor suite open source BI para proceder à implementação em ambiente real, visto ter sido a que obteve melhor pontuação final no método seguido. Esta

suite destacou-se da Pentaho nas funcionalidades que oferece (geolocalização, por exemplo), e na qualidade segundo o método Open BRR.

Após várias reuniões na empresa onde a implementação seria realizada, chegou-se a um conjunto de questões que deveriam ser respondidas por um sistema desta natureza, englobando quatro áreas de actuação: Recursos Humanos, Administração, Comercial e Logística. Foi assim desenhado o modelo dimensional da data warehouse que suportaria a implementação, e validado junto dos utilizadores.

Foi realizada a implementação na empresa MICROplásticos, SA, que passou pela instalação, pela automatização do ETL, e conseqüente configuração do sistema, por forma a disponibilizar a informação aos analistas e gestores, e assim ajudar no processo de decisão da empresa.

Verificou-se que a suite escolhida responde adequadamente às necessidades levantadas, possuindo as capacidades necessárias a uma empresa desta dimensão. Sendo um software open source, revelou-se uma excelente alternativa a soluções comerciais que pouco ou nada acrescentariam em termos de valor de negócio, e seriam certamente um factor de acréscimo de custos.

Assim, com este trabalho conseguimos atingir as principais contribuições a que nos propusemos:

- Mostrámos que as suites open source BI são uma alternativa a soluções comerciais para utilização em PME's, pois realizámos a implementação com sucesso respondendo a todas as questões que foram colocadas pelas diversas áreas da empresa onde intervimos.
- Realizámos um estudo comparativo das principais suites open source BI existentes no mercado.
- Apoiámos o processo de implementação de suites open source BI em contexto empresarial, pois foi neste contexto que realizámos a nossa implementação.
- Mostrámos que o open source é uma solução credível no contexto empresarial, pois implementámos uma ferramenta desta natureza nesse ambiente com resultados muito satisfatórios, e acolhida com entusiasmo.
- Incutimos e divulgámos a importância dos conceitos de BI, pois demos facilmente resposta a diversas questões que a empresa levaria muito tempo a obter, e ajudámos a obter conhecimento valioso para a própria empresa. Tal levou ainda a que se repensassem alguns processos internos.
- Melhorámos o processo de decisão numa PME intervindo em áreas específicas, pois, como já referido, demos resposta a várias questões que consumiam tempo e recursos.
- Contribuímos para o aumento da competitividade duma PME sem um aumento dos custos, pois ajudámos a que a empresa se conhecesse melhor e se torne mais competitiva, com uma solução totalmente open source que não levou a qualquer custo adicional.

- Divulgámos uma metodologia de avaliação de software open source, visto termos usado o Open BRR para avaliar e seleccionar a suite a implementar, e realizámos um artigo sobre o assunto que foi publicado numa conferência internacional.

Inserindo-se a empresa num universo de cerca de 98% do tecido empresarial, as PME's, e quase saindo já deste universo para o "lado" das grandes empresas, pode-se concluir que uma implementação BI com uma suite open source dará resposta às necessidades de uma fatia muito elevada do mercado das PME's.

A nível pessoal, considero que foi uma experiência muito satisfatória, e uma boa surpresa, a resposta dada por uma ferramenta desta natureza. Julgo que é uma vitória do open source e da sua comunidade de programadores, pois para as necessidades deste tipo de empresas não são necessárias soluções comerciais.

Uma das principais dificuldades encontradas neste trabalho foi conseguir pôr o Studio a comunicar com o Server da SpagoBI, nomeadamente com a versão do JDK, que só após alguma investigação (e a leitura o livro do Ogutu) foi resolvida. Outra dificuldade foi com os drivers do MySQL connector, que foi necessário actualizar para a versão mais recente para que a SpagoBI devolvesse resultados nos DataSets criados. Esta dificuldade só foi ultrapassada após pesquisa em fóruns. Na parte do ETL houve algumas dificuldades em obter de forma correcta os dados operacionais, pois a estrutura de tabelas do ERP usado é muito complexa, há muitas regras internas, e foi fundamental o apoio do departamento de informática da empresa para ultrapassar estas situações. Por fim, e como referido no capítulo da implementação, parte do ETL não foi realizado, pois não existem dados sobre o ponto de pessoal em nenhum formato digital que possam ser carregados, situação que será resolvida a breve prazo com alterações de procedimentos internos na empresa.

Como trabalho futuro, é premente continuar a acompanhar a evolução das suites open source BI. A portabilidade e o acesso à informação através de tablets e smartphones, o tratamento de enormes quantidades de dados, e a disponibilização da informação em tempo real são paradigmas e apostas que estão a surgir nas suites comerciais, e que seria importante analisar no mundo do open source. No mesmo sentido, a Cloud BI é uma solução que começa a ser adoptada e que desperta todo o interesse por parte das empresas, nomeadamente quanto ao custo de armazenamento e disponibilidade da informação, e que é premente acompanhar e evoluir.

O crescimento das comunidades e do número de ferramentas de software open source também requer um acompanhamento. Por fim, as metodologias de avaliação de software necessitam ser cimentadas e acompanhadas, pois de alguma forma ficaram estagnadas, como se referiu no estudo.

REFERÊNCIAS

- Amaral, L., Varajão, J., “Planeamento de Sistemas de Informação - 4ª Edição Actualizada e Aumentada”, FCA – Editora de Informática, 2007, pp. 6-9
- Antonelli, R., “Conhecendo o Business Intelligence”, CAP Accounting and Management, Revista Científica do TECAP, Nº3, Ano 3, Vol. 3, 2009, pp. 79-85
- Atos, <http://atos.net/en-us/home.html>
- Baptista, et al., “Open Source Software – Que oportunidades em Portugal”, APDSI - Associação para a Promoção e Desenvolvimento da Sociedade de Informação, 2004
- BSD, <http://www.linfo.org/bsdlicense.html> [acedido em Dezembro, 2013]
- Carlos, B., “Russia irá mudar para open source até ao final de 2014”, PortalOpenSource, 2011, <http://www.portalopensource.com/2011/01/17/russia-ira-mudar-para-open-source-ate-ao-final-de-2014/>
- Casos de sucesso, <http://www.esop.pt/casos-de-sucesso/> [acedido em Dezembro, 2013]
- Categories of free and nonfree software, 2002, <http://www.gnu.org/philosophy/categories.en.html> [acedido em Dezembro, 2013]
- CMU, <http://www.cmu.edu/silicon-valley/> [acedido em Dezembro, 2013]
- Davenport, T.; Prusak, L. Conhecimento empresarial. Rio de Janeiro: Campus, 1998
- Debian, http://www.debian.org/social_contract [acedido em Dezembro, 2013]
- Demarest, M., The politics of data warehousing, 1997
- Deprez, J., Alexandre, S., Comparing Assessment Methodologies for Free/Open Source Software: OpenBRR & QSOS, CETIC, Charleroi, Belgium, 2008
- Diário Coimbra, http://issuu.com/diariodecoimbra/docs/1000_maiores_2013 [acedido em Dezembro, 2013]
- Eclipse, <https://www.eclipse.org/> [acedido em Dezembro, 2013]

- Engineering, <http://www.spagoworld.org/xwiki/bin/view/SpagoWorld/Company> [acedido em Dezembro, 2013]
- Fayyad, U., Shapiro, G., Smyth, P., "From Data Mining to Knowledge Discovery in Databases", American Association for Artificial Intelligence, 1996
- Flores, "Que é Copyleft", 1988, <http://ramonflores.br.tripod.com/GNU/copyleft.html> [acedido em Dezembro, 2013]
- Free software, <https://www.gnu.org/philosophy/free-sw.html> [acedido em Dezembro, 2013]
- FSF, <http://www.fsf.org/> [acedido em Dezembro, 2013]
- Freeware, <http://www.techterms.com/definition/freeware> [acedido em Dezembro, 2013]
- Galliers, R. (Ed.), Information Analysis: Selected Readings, Addison-Wesley, 1987
- Global Alliance of SMEs, <http://www.globalsmes.org/html/index.php?func=expo2010&lan=en> [acedido em Dezembro, 2013]
- GPL, <https://www.gnu.org/copyleft/gpl.html> [acedido em Dezembro, 2013]
- Inmon, B., The Data Warehouse Budget, DM Review Magazine, 1997
- Inmon, W., "Building the Operational Data Store (2nd ed.)". New York: John Wiley & Sons, 1999
- Inmon, W., "Building the Data Warehouse (4rd ed.)", New York, John Wiley & Sons, Inc., 2005
- Intel, <http://www.intel.com> [acedido em Dezembro, 2013]
- Jaspersoft, <http://community.jaspersoft.com/> [acedido em Dezembro, 2013]
- JDK, <http://www.oracle.com/technetwork/java/javase/downloads/jdk7-downloads-1880260.html> [acedido em Dezembro, 2013]

- Kasabov, N. (2008). Evolving Intelligence in Humans and Machines: Integrative Evolving Connectionist Systems Approach. *IEEE Computational Intelligence*, Vol. 3, No., 3, pp. 23-37.
- Kimball, R.; Ross, M., "The Data Warehouse Toolkit (3rd ed.)", New York, John Wiley & Sons, Inc., 2013
- Kuhn, D., "Software Livre e as alterações no mercado de software no Brasil e no mundo – elementos para uma política governamental de software", Universidade Federal do Rio Grande do Sul, 2005
- Linden, K., "Dez regras clássicas para implementar Business Intelligence", Computerworld, <http://computerworld.uol.com.br/gestao/2011/01/11/dez-regras-classicas-para-implementar-business-intelligence/>, 2011 [acedido em Dezembro, 2013]
- Lista software open source,
http://pt.wikipedia.org/wiki/Anexo:Lista_de_softwares_open_source_para_Windows
[acedido em Dezembro, 2013]
- MDX , http://en.wikipedia.org/wiki/MultiDimensional_eXpressions [acedido em Dezembro, 2013]
- Metadata, <http://en.wikipedia.org/wiki/Metadata> [acedido em Dezembro, 2013]
- Multics, <http://web.mit.edu/multics-history/> [acedido em Dezembro, 2013]
- Negash, S., Gray, P., Business Intelligence, Ninth Americas Conference on Information System, Tampa, Florida, USA, 2003
- Negash, S., Business Intelligence, Communications of the Association for Information Systems, Volume13, 2004, 177-195, 2004
- Ogutu, S., SpagoBI 4.0 Baby Steps, The Ogutu Foundation, Nairobi, Kenia, 2012.
- Oliveira, R., "Opensource, oportunidade para as empresas!", 2012,
<http://formacao.rumos.pt/seccao3.rumos?id=75> [acedido em Dezembro, 2013]
- Open source, <http://opensource.org/about> [acedido em Dezembro, 2013]
- OpenBRR, <http://www.openbrr.org/> [acedido em Dezembro, 2013]
-

- OpenSuse, <http://software.opensuse.org/> [acedido em Dezembro, 2013]
- OSI, <http://opensource.org/osd-annotated> [acedido em Dezembro, 2013]
- Palo, <http://www.palo.net/> [acedido em Dezembro, 2013]
- Pentaho, <http://www.pentaho.com/> [acedido em Dezembro, 2013]
- Player, <http://www.vmware.com/products/player/> [acedido em Dezembro, 2013]
- Porter, M., Millar V., “How information gives you competitive advantage”, Harvard Business Review, July-August 1985
- Portugal e a realidade europeia, Semana Informática, 2012,
<http://www.semanainformatica.xl.pt/gest%C3%A3o/312-portugal-e-a-realidade-europeia.html> [acedido em Dezembro, 2013]
- Power, D.J. A Brief History of Decision Support Systems. DSSResources.COM,
<http://DSSResources.COM/history/dsshhistory.html>, version 4.0, March 10, 2007
[acedido em Dezembro, 2013]
- Primak, F., "Decisões com B.I. (Business Intelligence)", Rio de Janeiro: Ciência Moderna, 2008
- QSOS, http://www.qsos.org/?page_id=3 [acedido em Dezembro, 2013]
- Quinn, K., "Business Intelligence: conheça 4 piores práticas", Computerworld,
<http://computerworld.uol.com.br/gestao/2011/08/16/business-intelligence-conheca-4-piores-praticas/>, 2011[acedido em Dezembro, 2013]
- Raymond, E., The Cathedral and the Bazaar, Linux Kongress, Wurzburg, Germany, 1997
- Ribeiro, V., "O que é ETL?", Wordpress.com,
<http://vivaneribeiro1.wordpress.com/2011/06/28/o-que-e-etl-2/>, 2011[acedido em Dezembro, 2013]
- Schlegel, K., Sallam, R. L., Yuen, D., Tapadinhas, J., “Magic Quadrant for Business Intelligence and Analytics Platforms”, Gartner, February 2013

- Sezões, C., Oliveira, J., Baptista, M., “Business Intelligence”, SPI – Sociedade Portuguesa de Inovação, 2006,
http://www.bi4all.pt/docs/Manual_Business_Intelligence.pdf [acedido em Dezembro, 2013]
- Shareware, <http://www.techterms.com/definition/shareware> [acedido em Dezembro, 2013]
- Sinfic, “O Que é Uma Solução Open Source?”, Sinfic, SA, 2008
- Software Livre, http://pt.wikipedia.org/wiki/Software_livre [acedido em Dezembro, 2013]
- Software semi-livre, <https://www.gnu.org/philosophy/categories.en.html> [acedido em Dezembro, 2013]
- SpagoBI, <http://www.spagoworld.org/xwiki/bin/view/SpagoBI/Vision> [acedido em Dezembro, 2013]
- Stallman, R., The GNU Project, Open Sources, 1998
- Talend, <http://www.talend.com/> [acedido em Dezembro, 2013]
- TalendDocs,
<https://help.talend.com/display/TALENDOPENSTUDIOFORDATAINTEGRATIONUSERGUIDE52EN/Home> [acedido em Dezembro, 2013]
- Traffictool, <http://www.traffictool.net/vmware/opensuse123.html> [acedido em Dezembro, 2013]
- Tzu, S., ”A Arte da Guerra”, Texto Editores, Julho 2011
- Vanilla, <http://www.bpm-conseil.com/> [acedido em Dezembro, 2013]
- VMWare, <http://www.vmware.com/> [acedido em Dezembro, 2013]
- Watters, A., “Survey Finds Enterprise Is Overwhelmingly Open Source”, Readwrite, 2010, <http://readwrite.com/2010/08/10/survey-finds-enterprise-is-ove#awesm=~opFm40zJGJECsE> [acedido em Dezembro, 2013]

Wikimedia,
[http://commons.wikimedia.org/wiki/Category:Blank SVG maps of the world](http://commons.wikimedia.org/wiki/Category:Blank_SVG_maps_of_the_world) [acedido em Dezembro, 2013]

WikiStudio, https://wiki.spagobi.org/xwiki/bin/view/spagobi_studio/ [acedido em Dezembro, 2013]

Williams S., Williams N., “The Business Value of Business Intelligence”, Business Intelligence Journal, Fall 2003

X.org, <http://www.x.org/wiki/> [acedido em Dezembro, 2013]

Avaliação das métricas de cada categoria, do método Open BRR

O método Open BRR , tem uma escala de 1 a 5 (inaceitável a excelente) para cada uma das métricas referidas. Os critérios para a avaliação serão referidos em cada métrica. Esta análise foi efectuada em Maio de 2013, e todos os dados se reportam a essa altura.

Usabilidade

- **Experiência do interface para o utilizador final**

Descrição	Mede o quanto o interface é perceptível ao utilizador final (interface intuitivo, navegação, esquema de controle)				
Avaliação	5 - Excelente Simple e intuitivo, informação está bem organizada, não é necessário manual	4 - Bom	3 - Aceitável Leva algum tempo a aprender, informação pouco organizada, algum uso do manual	2 - Pobre	1 - Inaceitável Informação complexa, sem organização lógica, não pode ser usado sem manual
Pentaho	X				
SpagoBI	X				

- **Tempo de instalação dos pré-requisitos para a instalação do software open source**

Descrição	Tempo/esforço necessário para levantar o sistema, com todos os pré-requisitos instalados. Não inclui o sistema operativo				
Avaliação	5 - Excelente < 10 minutos	4 - Bom 10 - 30 minutos	3 - Aceitável 30 min - 1 hora	2 - Pobre 1 - 4 horas	1 - Inaceitável > 4 horas
Pentaho	X				
SpagoBI	X				

- **Tempo para instalação/configuração**

Descrição	Tempo para obter a <i>instant gratification</i> , mostra se o projecto está pensado para levar os utilizadores a funcionar o mais rapidamente possível				
Avaliação	5 - Excelente < 10 minutos	4 - Bom 10 - 30 minutos	3 - Aceitável 30 min - 1 hora	2 - Pobre 1 - 4 horas	1 - Inaceitável > 4 horas
Pentaho	X				
SpagoBI	X				

Qualidade

- **Nº de minor releases nos últimos 12 meses**

Descrição	Mede os updates planeados e a correcção de erros. Normalmente, os <i>service packs</i> nos produtos comerciais				
Avaliação	5 - Excelente	4 - Bom	3 - Aceitável	2 - Pobre	1 - Inaceitável
	2		1 ou 3		0 ou >3
Pentaho			X		
SpagoBI	X				

A Pentaho lançou a versão 4.8 a 29/11/2012, e tinha lançado a 4.5 a 20/04/2012

A Spago lançou a versão 3.5 a 04/07/2012 e a versão 3.6 a 06/11/2012.

- **Nº de point/patch releases nos últimos 12 meses**

Descrição	Normalmente são correcções para <i>bugs</i> P1 como <i>deadlocks</i> , memória e vulnerabilidades de segurança				
Avaliação	5 - Excelente	4 - Bom	3 - Aceitável	2 - Pobre	1 - Inaceitável
	3-4		1-2 ou 5-6		0 ou >6
Pentaho					X
SpagoBI			X		

Não há relato de actualizações da Pentaho nos últimos 12 meses.

A Spago lançou a versão 3.5.1 a 07/08/2012 depois de ter lançado a 3.5.0 a 04/07/2012

- **Nº de bugs relatados nos últimos 6 meses**

Descrição	Mede a qualidade do uso do projecto				
Avaliação	5 - Excelente	4 - Bom	3 - Aceitável	2 - Pobre	1 - Inaceitável
	< 50	50 - 100	100 - 500	500 - 1000	> 1000
Pentaho			X		
SpagoBI		X			

Consultas efectuadas ao Jira de ambas as plataformas em

<http://jira.pentaho.com/issues/?jql=project%20%3D%20BISERVER%20AND%20issue%20type%20%3D%20Bug%20AND%20created%20%3E%3D%202012-12-07%20AND%20created%20%3C%3D%202013-05-07>

Resultado: 402

<http://www.spagoworld.org/jira/issues/?jql=project%20%3D%20SPAGOBI%20AND%20issuetype%20%3D%20Bug%20AND%20created%20%3E%3D%202012-12-07%20AND%20created%20%3C%3D%202013-05-07>

Resultado: 50

- **Nº de bugs resolvidos nos últimos 6 meses (comparado com o nº de bugs relatados)**

Descrição	Mede a rapidez de resolução de <i>bugs</i>				
Avaliação	5 - Excelente	4 - Bom	3 - Aceitável	2 - Pobre	1 – Inaceitável
	> 75%	60% - 75%	45% - 60%	25% - 45%	< 25%
Pentaho		X			
SpagoBI				X	

[http://jira.pentaho.com/issues/?jql=project%20%3D%20BISERVER%20AND%20issue%20type%20%3D%20Bug%20AND%20status%20in%20\(Resolved%2C%20Closed\)%20AND%20created%20%3E%3D%202012-12-07%20AND%20created%20%3C%3D%202013-05-07](http://jira.pentaho.com/issues/?jql=project%20%3D%20BISERVER%20AND%20issue%20type%20%3D%20Bug%20AND%20status%20in%20(Resolved%2C%20Closed)%20AND%20created%20%3E%3D%202012-12-07%20AND%20created%20%3C%3D%202013-05-07)

Resultado: 286 -> 71%

[http://www.spagoworld.org/jira/issues/?jql=project%20%3D%20SPAGOBI%20AND%20issuetype%20%3D%20Bug%20AND%20status%20in%20\(Resolved%2C%20Closed\)%20AND%20created%20%3E%3D%202012-12-07%20AND%20created%20%3C%3D%202013-05-07](http://www.spagoworld.org/jira/issues/?jql=project%20%3D%20SPAGOBI%20AND%20issuetype%20%3D%20Bug%20AND%20status%20in%20(Resolved%2C%20Closed)%20AND%20created%20%3E%3D%202012-12-07%20AND%20created%20%3C%3D%202013-05-07)

[http://www.spagoworld.org/jira/issues/?jql=project%20%3D%20SPAGOBI%20AND%20issuetype%20%3D%20Bug%20AND%20status%20in%20\(Resolved%2C%20Closed\)%20AND%20created%20%3E%3D%202012-12-07%20AND%20created%20%3C%3D%202013-05-07](http://www.spagoworld.org/jira/issues/?jql=project%20%3D%20SPAGOBI%20AND%20issuetype%20%3D%20Bug%20AND%20status%20in%20(Resolved%2C%20Closed)%20AND%20created%20%3E%3D%202012-12-07%20AND%20created%20%3C%3D%202013-05-07)

Resultado: 15 -> 30%

- **Nº de P1/critical bugs relatados**

Descrição	Mede a gravidade dos problemas de qualidade encontrados				
Avaliação	5 – Excelente	4 - Bom	3 - Aceitável	2 – Pobre	1 – Inaceitável
	0	1 - 5	5 - 10	10 - 20	> 20
Pentaho					X
SpagoBI					X

[http://jira.pentaho.com/issues/?jql=project%20%3D%20BISERVER%20AND%20issue%20type%20%3D%20Bug%20AND%20created%20%3E%3D%202012-12-07%20AND%20created%20%3C%3D%202013-05-07%20AND%20priority%20in%20\(Blocker%2C%20Critical%2C%20Severe\)](http://jira.pentaho.com/issues/?jql=project%20%3D%20BISERVER%20AND%20issue%20type%20%3D%20Bug%20AND%20created%20%3E%3D%202012-12-07%20AND%20created%20%3C%3D%202013-05-07%20AND%20priority%20in%20(Blocker%2C%20Critical%2C%20Severe))

[http://jira.pentaho.com/issues/?jql=project%20%3D%20BISERVER%20AND%20issue%20type%20%3D%20Bug%20AND%20created%20%3E%3D%202012-12-07%20AND%20created%20%3C%3D%202013-05-07%20AND%20priority%20in%20\(Blocker%2C%20Critical%2C%20Severe\)](http://jira.pentaho.com/issues/?jql=project%20%3D%20BISERVER%20AND%20issue%20type%20%3D%20Bug%20AND%20created%20%3E%3D%202012-12-07%20AND%20created%20%3C%3D%202013-05-07%20AND%20priority%20in%20(Blocker%2C%20Critical%2C%20Severe))

[http://jira.pentaho.com/issues/?jql=project%20%3D%20BISERVER%20AND%20issue%20type%20%3D%20Bug%20AND%20created%20%3E%3D%202012-12-07%20AND%20created%20%3C%3D%202013-05-07%20AND%20priority%20in%20\(Blocker%2C%20Critical%2C%20Severe\)](http://jira.pentaho.com/issues/?jql=project%20%3D%20BISERVER%20AND%20issue%20type%20%3D%20Bug%20AND%20created%20%3E%3D%202012-12-07%20AND%20created%20%3C%3D%202013-05-07%20AND%20priority%20in%20(Blocker%2C%20Critical%2C%20Severe))

Resultado: 304

[http://www.spagoworld.org/jira/issues/?jql=project%20%3D%20SPAGOBI%20AND%20issuetype%20%3D%20Bug%20AND%20priority%20in%20\(Blocker%2C%20Critical%2C%20Major\)%20AND%20created%20%3E%3D%202012-12-07%20AND%20created%20%3C%3D%202013-05-07%20ORDER%20BY%20priority%20DESC](http://www.spagoworld.org/jira/issues/?jql=project%20%3D%20SPAGOBI%20AND%20issuetype%20%3D%20Bug%20AND%20priority%20in%20(Blocker%2C%20Critical%2C%20Major)%20AND%20created%20%3E%3D%202012-12-07%20AND%20created%20%3C%3D%202013-05-07%20ORDER%20BY%20priority%20DESC)

[http://www.spagoworld.org/jira/issues/?jql=project%20%3D%20SPAGOBI%20AND%20issuetype%20%3D%20Bug%20AND%20priority%20in%20\(Blocker%2C%20Critical%2C%20Major\)%20AND%20created%20%3E%3D%202012-12-07%20AND%20created%20%3C%3D%202013-05-07%20ORDER%20BY%20priority%20DESC](http://www.spagoworld.org/jira/issues/?jql=project%20%3D%20SPAGOBI%20AND%20issuetype%20%3D%20Bug%20AND%20priority%20in%20(Blocker%2C%20Critical%2C%20Major)%20AND%20created%20%3E%3D%202012-12-07%20AND%20created%20%3C%3D%202013-05-07%20ORDER%20BY%20priority%20DESC)

[http://www.spagoworld.org/jira/issues/?jql=project%20%3D%20SPAGOBI%20AND%20issuetype%20%3D%20Bug%20AND%20priority%20in%20\(Blocker%2C%20Critical%2C%20Major\)%20AND%20created%20%3E%3D%202012-12-07%20AND%20created%20%3C%3D%202013-05-07%20ORDER%20BY%20priority%20DESC](http://www.spagoworld.org/jira/issues/?jql=project%20%3D%20SPAGOBI%20AND%20issuetype%20%3D%20Bug%20AND%20priority%20in%20(Blocker%2C%20Critical%2C%20Major)%20AND%20created%20%3E%3D%202012-12-07%20AND%20created%20%3C%3D%202013-05-07%20ORDER%20BY%20priority%20DESC)

Resultado 36

- **Tempo médio de bugs P1 nos últimos 6 meses**

Descrição	Mede a capacidade de resposta a resolução de problemas críticos				
Avaliação	5 – Excelente	4 - Bom	3 - Aceitável	2 – Pobre	1 – Inaceitável

	<1 semana	1 - 2 semanas	2 - 3 semanas	3 - 4 semanas	> 4 semanas
Pentaho					X
SpagoBI	X				

[http://jira.pentaho.com/issues/?jql=project%20%3D%20BISERVER%20AND%20issue%20type%20%3D%20Bug%20AND%20status%20in%20\(Resolved%2C%20Closed\)%20AND%20created%20%3E%3D%202012-12-07%20AND%20created%20%3C%3D%202013-05-07%20AND%20priority%20in%20\(Blocker%2C%20Critical%2C%20Severe\)%20ORDER%20BY%20remainingEstimate%20ASC%2C%20due%20ASC](http://jira.pentaho.com/issues/?jql=project%20%3D%20BISERVER%20AND%20issue%20type%20%3D%20Bug%20AND%20status%20in%20(Resolved%2C%20Closed)%20AND%20created%20%3E%3D%202012-12-07%20AND%20created%20%3C%3D%202013-05-07%20AND%20priority%20in%20(Blocker%2C%20Critical%2C%20Severe)%20ORDER%20BY%20remainingEstimate%20ASC%2C%20due%20ASC)

Resultado: 34 dias

[http://www.spagoworld.org/jira/issues/?jql=project%20%3D%20SPAGOBI%20AND%20issuetype%20%3D%20Bug%20AND%20priority%20in%20\(Blocker%2C%20Critical%2C%20Major\)%20AND%20status%20in%20\(Resolved%2C%20Closed\)%20AND%20created%20%3E%3D%202012-12-07%20AND%20created%20%3C%3D%202013-05-07%20ORDER%20BY%20priority%20DESC](http://www.spagoworld.org/jira/issues/?jql=project%20%3D%20SPAGOBI%20AND%20issuetype%20%3D%20Bug%20AND%20priority%20in%20(Blocker%2C%20Critical%2C%20Major)%20AND%20status%20in%20(Resolved%2C%20Closed)%20AND%20created%20%3E%3D%202012-12-07%20AND%20created%20%3C%3D%202013-05-07%20ORDER%20BY%20priority%20DESC)

Resultado: 1,8 dias

Segurança

- **Número de vulnerabilidades de segurança nos últimos seis meses que são moderadamente a extremamente críticas**

Descrição	Mede a qualidade relativamente a vulnerabilidades de segurança. Quanto susceptível está o software a vulnerabilidades de segurança				
	5 - Excelente	4 - Bom	3 - Aceitável	2 - Pobre	1 - Inaceitável
	0	1 - 2	3 - 4	5 - 6	> 6
Pentaho	X				
SpagoBI	X				

[http://jira.pentaho.com/issues/?jql=project%20%3D%20BISERVER%20AND%20issue%20type%20in%20standardIssueTypes\(\)%20AND%20priority%20in%20\(Blocker%2C%20Critical%2C%20Severe%2C%20Medium\)%20AND%20created%20%3E%3D%202012-12-07%20AND%20created%20%3C%3D%202013-05-07%20AND%20text%20~%20security%20ORDER%20BY%20issuetype%20DESC%2C%20resolution%20ASC%2C%20remainingEstimate%20ASC%2C%20due%20ASC](http://jira.pentaho.com/issues/?jql=project%20%3D%20BISERVER%20AND%20issue%20type%20in%20standardIssueTypes()%20AND%20priority%20in%20(Blocker%2C%20Critical%2C%20Severe%2C%20Medium)%20AND%20created%20%3E%3D%202012-12-07%20AND%20created%20%3C%3D%202013-05-07%20AND%20text%20~%20security%20ORDER%20BY%20issuetype%20DESC%2C%20resolution%20ASC%2C%20remainingEstimate%20ASC%2C%20due%20ASC)

Resultado: 0

[http://www.spagoworld.org/jira/issues/?jql=project%20%3D%20SPAGOBI%20AND%20issuetype%20in%20standardIssueTypes\(\)%20AND%20priority%20in%20\(Blocker%2C%20Critical%2C%20Major\)%20AND%20status%20in%20\(Resolved%2C%20Closed\)%20AND%20created%20%3E%3D%202012-12-07%20AND%20created%20%3C%3D%202013-05-07%20ORDER%20BY%20priority%20DESC](http://www.spagoworld.org/jira/issues/?jql=project%20%3D%20SPAGOBI%20AND%20issuetype%20in%20standardIssueTypes()%20AND%20priority%20in%20(Blocker%2C%20Critical%2C%20Major)%20AND%20status%20in%20(Resolved%2C%20Closed)%20AND%20created%20%3E%3D%202012-12-07%20AND%20created%20%3C%3D%202013-05-07%20ORDER%20BY%20priority%20DESC)

[07%20AND%20created%20%3C%3D%202013-05-](#)

[07%20AND%20text%20~%20security%20ORDER%20BY%20priority%20DESC](#)

Resultado: 0

- **Número de vulnerabilidades de segurança ainda abertas (sem correcção)**

Descrição	Mede a capacidade que o projecto tem em resolver todas as vulnerabilidades de segurança				
Avaliação	5 – Excelente	4 - Bom	3 – Aceitável	2 – Pobre	1 – Inaceitável
	0	1	2	3 - 5	> 5
Pentaho	X				
SpagoBI	X				

- **Há informação dedicada para segurança (página web, wiki, etc)?**

Descrição	Mede o conhecimento e quanto são levados a sério os problemas de segurança				
Avaliação	5 – Excelente	4 - Bom	3 - Aceitável	2 – Pobre	1 – Inaceitável
	Sim, bem mantida		Sim		Não
Pentaho	X				
SpagoBI	X				

<http://wiki.pentaho.com/display/ServerDoc2x/Security>

http://wiki.spagobi.org/xwiki/bin/view/spagobi_server/User+Management_v3

Desempenho

- **Há testes de desempenho e relatórios de referência (Benchmark) disponíveis?**

Descrição	Mede se há testes de desempenho e benchmarks publicados – tipicamente em comparação com soluções equivalentes				
Avaliação	5 – Excelente	4 - Bom	3 - Aceitável	2 – Pobre	1 – Inaceitável
	Sim, com bons resultados		Sim		Não
Pentaho					X
SpagoBI					X

- **Optimização de desempenho e configuração**

Descrição	Mede se há alguma documentação ou ferramenta que ajude a afinar o desempenho (informação sobre CPU, disco, rede...)				
Avaliação	5 – Excelente	4 - Bom	3 - Aceitável	2 – Pobre	1 – Inaceitável
	Sim, extensa		Sim, alguma		Não

Pentaho			X		
SpagoBI					X

<http://wiki.pentaho.com/display/EAI/Performance+tuning>

Escalabilidade

- **Referência de implementação**

Descrição	Mede o quanto o software é escalável e testado em ambiente real, através de implementação no mundo real				
Avaliação	5 – Excelente Sim, com publicação do nº de utilizadores	4 - Bom	3 - Aceitável Sim	2 – Pobre	1 – Inaceitável Não
Pentaho	X				
SpagoBI	X				

- **Projectado para escalabilidade**

Descrição	Mede se o projecto foi concebido com a escalabilidade em mente. É thread-safe? Pode ser executado em ambiente de cluster? Pode o hardware resolver problemas de desempenho?				
Avaliação	5 – Excelente Sim, extensa	4 – Bom	3 - Aceitável Sim, alguma	2 – Pobre	1 – Inaceitável Não
Pentaho	X				
SpagoBI	X				

Arquitectura

- **Há plug-ins de terceiros?**

Descrição	Mede se o projecto foi desenhado para extensibilidade por terceiros				
Avaliação	5 – Excelente > 10	4 – Bom 6 - 10	3 - Aceitável 2 - 5	2 – Pobre 1	1 – Inaceitável 0
Pentaho	X				
SpagoBI					X

- **API pública / Serviço Externo**

Descrição	Permite extensões via API pública, também permite personalização?				
Avaliação	5 – Excelente Sim, extensa	4 – Bom	3 - Aceitável Sim	2 – Pobre	1 – Inaceitável Não
Pentaho	X				
SpagoBI	X				

Apoio

- **Volume médio de listas de discussão gerais nos últimos 6 meses**

Descrição	As listas de discussão gerais são o primeiro sítio onde as pessoas vão para ajuda gratuita				
Avaliação	5 – Excelente	4 – Bom	3 - Aceitável	2 – Pobre	1 – Inaceitável
	> 720 mensagens por mês	300 - 720 msg por mês	150 - 300 msg por mês	30 - 150 msg por mês	< 30 msg por mês
Pentaho			X		
SpagoBI				X	

- **Qualidade do apoio profissional**

Descrição	Apoio profissional que ajuda a afinar a implementação e resolução de problemas é sempre desejável				
Avaliação	5 – Excelente	4 – Bom	3 - Aceitável	2 – Pobre	1 – Inaceitável
	Instalação + Resolução de problemas + integração / personalização do apoio		Apenas apoio à instalação		Sem apoio profissional
Pentaho	X				
SpagoBI	X				

Documentação

- **Existência de vários tipos de documentação**

Descrição	Uma boa ferramenta de documentação deve incluir documentação para vários grupos de utilizadores em vários formatos				
Avaliação	5 – Excelente	4 – Bom	3 - Aceitável	2 – Pobre	1 – Inaceitável
	Instalar / implementar, utilizadores, administração, optimização, actualização, documentação de desenvolvimento está disponível em múltiplos formatos (pdf, html)	Instalar / implementar, utilizadores, administração, manuais de melhoramento disponíveis em vários formatos	Instalar / implementar e manual de utilizador disponível	Apenas existe documentação de instalação baseada em texto	Sem documentação. Um ficheiro README não conta
Pentaho	X				
SpagoBI	X				

- **Contribuição do utilizador**

Descrição	As melhores ajudas vêm muitas vezes de exemplos fornecidos por utilizadores. É o feedback de pessoas que usaram o produto				
Avaliação	5 – Excelente As pessoas podem contribuir, e as contribuições são editadas / filtradas por especialistas	4 – Bom	3 - Aceitável As pessoas podem contribuir	2 – Pobre	1 – Inaceitável Utilizadores não podem contribuir
Pentaho	X				
SpagoBI	X				

Adopção

- **Quantos títulos de livros na Amazon.com devolve a consulta: "subject:computer and title:component name"**

Descrição	A disponibilidade de livros é certamente bom. Uma forma padronizada de contagem dos livros disponíveis é importante				
Avaliação	5 – Excelente > 15	4 – Bom 6 - 15	3 - Aceitável 3 - 6	2 – Pobre 1 - 3	1 – Inaceitável 0
Pentaho		X			
SpagoBI				X	

- **Referência de implementação**

Descrição	Mede o quanto o software é escalável e testado em ambiente real, através de implementação no mundo real				
Avaliação	5 – Excelente Sim, com publicação do nº de utilizadores	4 – Bom	3 - Aceitável Sim	2 – Pobre	1 – Inaceitável Não
Pentaho	X				
SpagoBI	X				

Comunidade

- **Volume médio de listas de discussão geral nos últimos 6 meses**

Descrição	As listas de discussão gerais são o local onde a comunidade se interajuda				
Avaliação	5 – Excelente	4 – Bom	3 - Aceitável	2 – Pobre	1 – Inaceitável

	> 720 mensagens por mês	300 - 720 msg por mês	150 - 300 msg por mês	30 - 150 msg por mês	< 30 msg por mês
Pentaho			X		
SpagoBI			X		

- **Nº de contribuintes de código nos últimos 6 meses**

Descrição	Os contribuintes de código normalmente promovem a construção de uma comunidade à volta do projecto. Quanto maior o número de contribuintes de código, melhor o apoio da comunidade				
Avaliação	5 – Excelente > 50	4 – Bom 20 - 50	3 - Aceitável 10 - 20	2 – Pobre 5 - 10	1 – Inaceitável < 5
Pentaho	X				
SpagoBI	X				

Em ambos os casos, nos últimos 6 meses não foi possível analisar, esta indicação é pela lista de contribuintes de código disponibilizada pelas plataformas

Profissionalismo

- **Driver do projecto**

Descrição	O driver do projecto realiza a gestão de projectos, angariação de dinheiro, etc.				
Avaliação	5 – Excelente Fundação independente apoiada por corporações (estilo Apache / OSDL)	4 – Bom Corporativo (estilo MySQL)	3 - Aceitável	2 – Pobre Grupos	1 – Inaceitável Individuais
Pentaho		X			
SpagoBI		X			

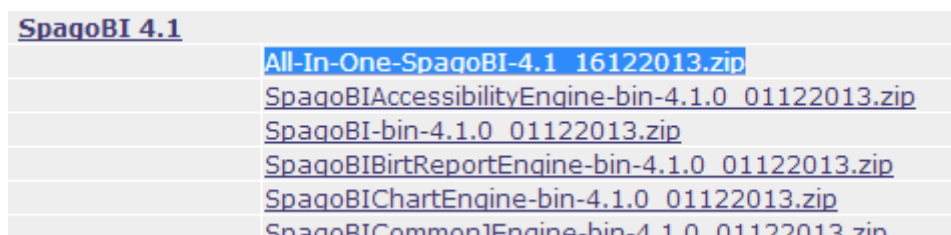
- **Dificuldade de pertencer à equipa de desenvolvimento**

Descrição	Para garantir a qualidade do software, projectos maduros devem ser selectivos na aceitação de membros. Novos projectos não têm opção				
Avaliação	5 – Excelente Só depois de ser um contribuidor activo externo por algum tempo	4 – Bom	3 - Aceitável Bastante difícil, deve contribuir com correcções aceites por algum tempo	2 – Pobre	1 – Inaceitável Qualquer um pode entrar
Pentaho			X		
SpagoBI			X		

Instalação da suite open source SpagoBI, em Linux

A versão mais recente da suite SpagoBI é a 4.1.

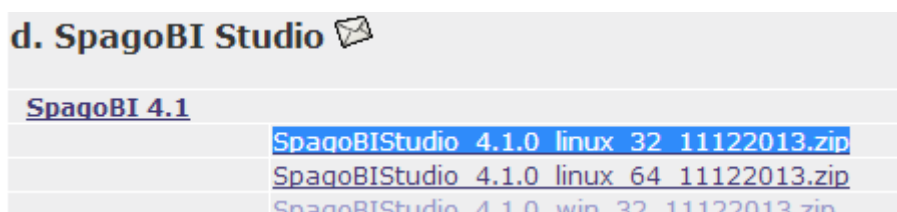
Para fazer download e posterior instalação é necessário ir ao link http://forge.ow2.org/project/showfiles.php?group_id=204 e clicar em All-In-One-SpagoBI-4.1_16122013.zip, como demonstra a figura 1, abaixo.



SpagoBI 4.1	
	All-In-One-SpagoBI-4.1_16122013.zip
	SpagoBIAccessibilityEngine-bin-4.1.0_01122013.zip
	SpagoBI-bin-4.1.0_01122013.zip
	SpagoBIBirtReportEngine-bin-4.1.0_01122013.zip
	SpagoBIChartEngine-bin-4.1.0_01122013.zip
	SpagoBICommonEngine-bin-4.1.0_01122013.zip

Figura 1. Link do download do SpagoBI All in One 4.1

Para criar reports BIRT (Business Intelligence and Reporting Tools), documentos OLAP, gráficos, cockpits, etc., é necessário usar o SpagoBI Studio, pelo que aproveitámos e fizemos logo o seu download na mesma página, clicando em [SpagoBIStudio 4.1.0 linux 32 11122013.zip](#), como demonstra a figura 2.




d. SpagoBI Studio 	
SpagoBI 4.1	
	SpagoBIStudio 4.1.0 linux 32 11122013.zip
	SpagoBIStudio 4.1.0 linux 64 11122013.zip
	SpagoBIStudio 4.1.0 win 32 11122013.zip

Figura 2. Link do download do SpagoBI Studio 4.1

Adicionalmente é necessário instalar uma versão do JDK (Java Development Kit), e definir a variável de sistema JAVA_HOME.

Como a imagem do OpenSuse instalada contém já muitas ferramentas pré-instaladas, podemos, com o “Dolphin”, ir a “Downloads”, clicar sobre o ficheiro All-In-One-SpagoBI-4.1_16122013.zip e extrair o seu conteúdo para home/spagoBI (opção, no caso).

Para instalar o JDK, fizemos o download da versão 6, pois (Oguto, 2012) refere problemas, que verificámos, com a versão 7 do JDK. Como não é a versão mais recente, pode encontrar-se em *Previous Releases*, no site da Oracle (JDK, 2014), como mostra a figura 3.

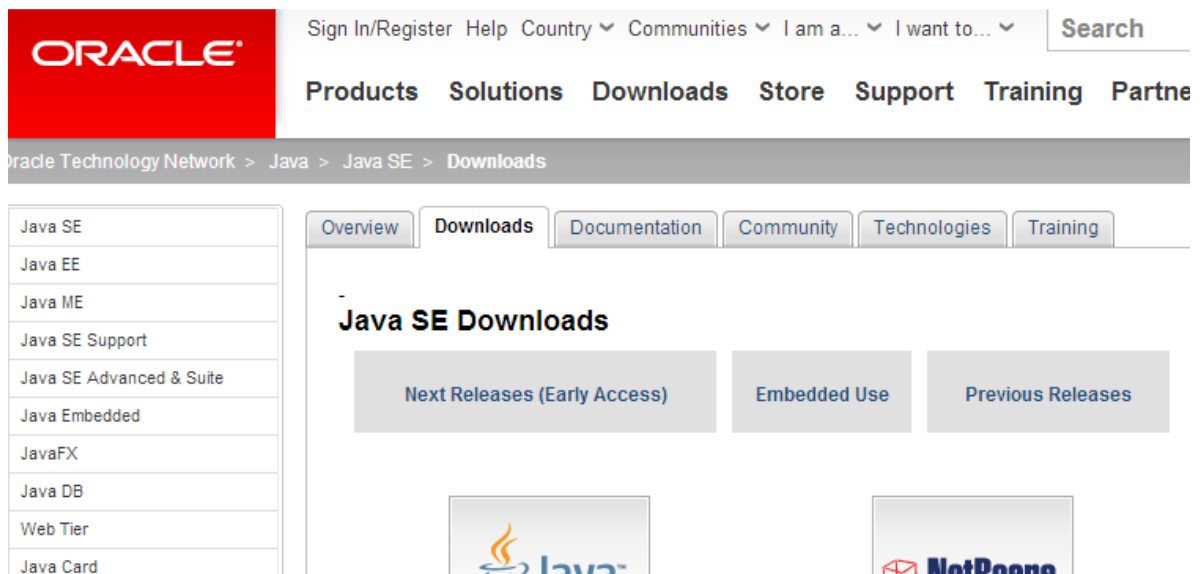


Figura 3. Página da Oracle onde estão as versões anteriores do JDK

A figura 4 mostra que fizemos o download para a pasta home/spagoBI.

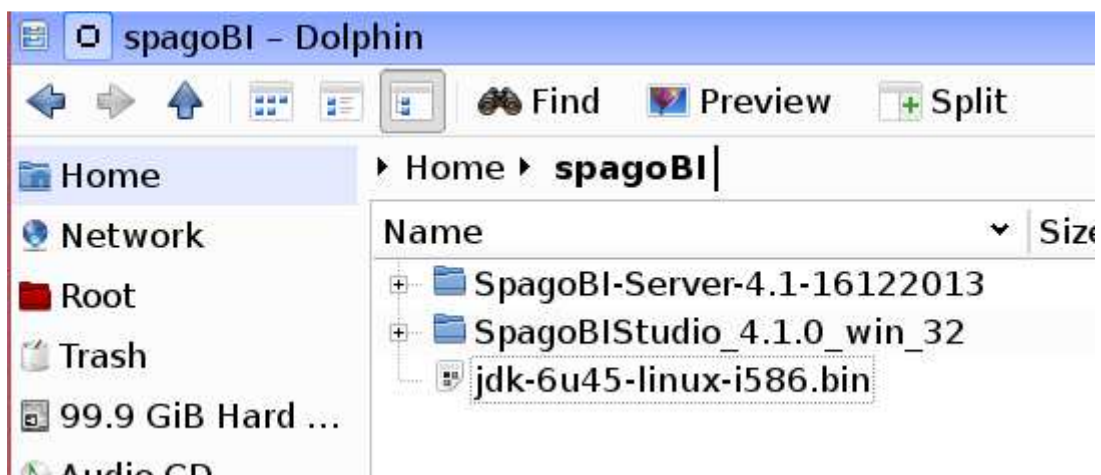


Figura 4. Pasta do download para instalação do JDK

De seguida, abrindo a consola, através do “Terminal”, foram dados os passos mostrados na figura 5 para instalar o JDK.

```
user@linux-x2pw:~> cd spagoBI/
user@linux-x2pw:~/spagoBI> chmod 700 jdk-6u45-linux-i586.bin
user@linux-x2pw:~/spagoBI> sudo ./jdk-6u45-linux-i586.bin
```

Figura 5. Passos para a instalação do JDK

A figura 6 mostra os passos para abrir o editor “vi” para poder definir a variável JAVA_HOME.

```

user@linux-x2pw:~/spagoBI> ls
jdk1.6.0_45          SpagoBI-Server-4.1-16122013
jdk-6u45-linux-i586.bin SpagoBIStudio_4.1.0_win_32
user@linux-x2pw:~/spagoBI> cd jdk1.6.0_45/
user@linux-x2pw:~/spagoBI/jdk1.6.0_45> ls
bin      db      jre  LICENSE  README.html  THIRDPARTYLICENSEREADME.txt
COPYRIGHT include lib man      src.zip
user@linux-x2pw:~/spagoBI/jdk1.6.0_45> pwd
/home/user/spagoBI/jdk1.6.0_45
user@linux-x2pw:~/spagoBI/jdk1.6.0_45> vi ~/.bashrc █

```

Figura 6. Passos para definir a JAVA_HOME

No editor, devem ser inseridos os comandos mostrados na figura 7.

```

export JAVA_HOME=/home/user/spagoBI/jdk1.6.0_45
export PATH=$JAVA_HOME/bin:$PATH

```

```
:wq█
```

Figura 7. Comandos de definição da variável JAVA_HOME

Para gravar e sair do editor, pressionar “Esc” e escrever “:wq”. A figura 8 mostra os passos seguintes.

```

user@linux-x2pw:~/spagoBI/jdk1.6.0_45> source ~/.bashrc
user@linux-x2pw:~/spagoBI/jdk1.6.0_45> which java
/home/user/spagoBI/jdk1.6.0_45/bin/java
user@linux-x2pw:~/spagoBI/jdk1.6.0_45> █

```

Figura 8. Comandos para verificar o caminho exacto do executável java

Para verificar a versão do JDK, instalada, pode ser executado o comando da figura 9.

```

user@linux-x2pw:~/spagoBI/jdk1.6.0_45> java -version
java version "1.6.0_45"
Java(TM) SE Runtime Environment (build 1.6.0_45-b06)
Java HotSpot(TM) Client VM (build 20.45-b01, mixed mode, sharing)
user@linux-x2pw:~/spagoBI/jdk1.6.0_45> █

```

Figura 9. Comando para verificar a versão do JDK instalada

Para trabalhar com o SpagoBI, devemos executar os comandos da figura 10 que vão arrancar o servidor SpagoBI.

```

user@linux-x2pw:~/spagoBI> cd jdk1.6.0_45/
user@linux-x2pw:~/spagoBI/jdk1.6.0_45> cd..
user@linux-x2pw:~/spagoBI> cd SpagoBI-Server-4.1-16122013/
user@linux-x2pw:~/spagoBI/SpagoBI-Server-4.1-16122013> cd bin/
user@linux-x2pw:~/spagoBI/SpagoBI-Server-4.1-16122013/bin> chmod +x *.sh
user@linux-x2pw:~/spagoBI/SpagoBI-Server-4.1-16122013/bin> ./startup.sh █

```

Figura 10. Comandos para arrancar o servidor SpagoBI

Por fim, estamos prontos a explorar a ferramenta, abrindo o browser através do endereço <http://localhost:8080/SpagoBI>, e fazendo login, como mostra a figura 11.



Figura 11. SpagoBI no browser, preparado para fazer login

Anexo C

Artigo “Analysis of Open Source Business Intelligence Suites”

8ª Conferência Ibérica de Sistemas e Tecnologias de Informação - CISTI'2013
19-22 de Junho de 2013, Lisboa - Portugal

Analysis of Open Source Business Intelligence Suites

Análise de Suites Open Source Business Intelligence

António Marinheiro

Instituto Politécnico de Coimbra
ISEC - Instituto Superior de Engenharia de Coimbra
Coimbra, Portugal
a21200015@alunos.isec.pt

Jorge Bernardino

Instituto Politécnico de Coimbra
ISEC - Instituto Superior de Engenharia de Coimbra
Coimbra, Portugal
jorge@isec.pt

Abstract— In recent time, technology applications in different fields, especially Business Intelligence (BI) have been developed rapidly and considered to be one of the most significant uses of information technology. BI is a broad category of applications and technologies for gathering, storing, analyzing, and providing access to data to help enterprise users make better business decisions. Whereas in the past the Business Intelligence market was strictly dominated by closed source and commercial tools, the last few years were characterized by the birth of open source solutions. This represents a tremendous competitive advantage, however the choice of a suitable open source BI suite is a challenge. The present study evaluated and compared the last versions of the five main Open Source Business Intelligence suites: JasperSoft, Palo, Pentaho, SpagoBI and Vanilla.

Keywords: Business Intelligence (BI); Free and Libre Open Source Software (FLOSS); Open Source Business Intelligence suites

Resumo—Nos últimos anos, as aplicações tecnológicas em diferentes áreas, especialmente de Business Intelligence (BI) desenvolveram-se rapidamente e são consideradas como sendo das de uso mais significativo nas tecnologias da informação. Business Intelligence é uma categoria ampla de aplicações e tecnologias para recolha, armazenamento, análise e disponibilização de acesso a dados para ajudar os utilizadores nas empresas a tomarem melhores decisões de negócio. Considerando que, no passado, o mercado de Business Intelligence foi estritamente dominado por ferramentas comerciais, os tempos recentes foram caracterizados pelo nascimento de soluções open source. Isto representa uma enorme vantagem competitiva, no entanto, a escolha de uma Suite BI open source é um grande desafio. O presente estudo avaliou e comparou as últimas versões das cinco principais suites de Business Intelligence open source: Jaspersoft, Palo, Pentaho, SpagoBI e Vanilla.

Palavras-chave: Business Intelligence (BI); Free and Libre Open Source Software (FLOSS); Suites Business Intelligence Open Source

I. Introdução

No ambiente empresarial actual, em virtude das grandes alterações que se têm verificado, as organizações necessitam ser suficientemente ágeis para antecipar as mudanças, encontrando soluções e adaptando-se, ou criando bases para vantagens competitivas face à concorrência. Neste contexto, há cada vez mais a necessidade das organizações terem informações relevantes que auxiliem a tomada de decisão e de ferramentas para servir esses propósitos. Desde os primórdios da estratégia organizacional, de Sun Tzu (*A Arte da Guerra*, escrito há cerca de 2500 anos) [1], que existe uma preocupação na forma de se conhecerem os pontos fortes e fracos, assim como os dos concorrentes. Mais recentemente, Michael Porter refere que em qualquer empresa a informação tecnológica tem um efeito poderoso na vantagem competitiva em termos de custos e diferenciação [2]. Antes da informatização da informação, não sendo esta estruturada, as decisões tomadas tinham sempre um grande grau de intuição. Com a complexidade tecnológica e concorrencial dos dias de hoje, as empresas estão obrigadas a prestar uma atenção redobrada ao processo de controlo interno, de desempenho e aos sistemas de suporte à decisão.

Neste âmbito, Williams [3] afirma que a engenharia de processos e a análise da mudança estabelecem as bases para incrementar o valor do negócio.

O Business Intelligence (BI) surge assim como um conjunto de conceitos e métodos para auxiliar e aumentar a capacidade de tomada de decisão nas organizações, transformando os seus dados em conhecimento [4]. Esta capacidade só é possível com o auxílio de ferramentas de BI, independentes entre si, ou reunidas numa única plataforma, como as estudadas neste artigo.

O termo BI surgiu pela primeira vez em 1958, definido por H. P. Luhn num artigo publicado no IBM Journal [5]. Este termo foi retomado e redefinido por Howard Dresner da Gartner Group em 1989, como sendo um conjunto de conceitos e métodos para melhorar a tomada de decisão empresarial recorrendo a sistemas de apoio baseados em factos [4]. Nesse tempo, a indústria de software falava em acrónimos como Sistemas de Apoio à Decisão (DSS – Decision Support Systems) e Sistemas de Informação Executiva (EIS – Enterprise Information Systems), e Dresner procurava um conceito que não colocasse apenas o foco do lado tecnológico, mas principalmente do lado do negócio.

Deste modo o BI tem como principal objectivo estabelecer a ponte entre os dados gerados no dia-a-dia dos negócios das empresas, transformando-os em informação útil, através de contextualização, categorização, cálculo, correção e agregação dos dados [6]. Esta informação, através da intervenção humana a nível de comparação, análise de consequências, relacionamento e reflexão, vai traduzir-se em conhecimento. O suporte à decisão é dado pelo maior número possível de informações e conhecimento disponíveis, para que sejam escolhidas algumas ou apenas uma entre as várias

alternativas para as acções a serem realizadas [7]. Os dados, a informação e o conhecimento são assim a vantagem competitiva para as empresas, que a auxiliam nas decisões a tomar para se diferenciarem e ganhar vantagem competitiva face à concorrência.

Por exemplo, Sezões [8] afirma que os sistemas de BI têm em comum um conjunto de importantes objectivos fundamentais:

- Acesso a dados fiáveis – a fiabilidade dos dados, a sua integração e compreensão entre diversas áreas é fundamental a um consciente exercício de gestão;
- Aumento da transparência e compreensão do negócio – a disponibilização de conhecimento em tempo real (“o quê?”, “o quanto?”, “o quando?”, “o onde?” e “o como?”) permite aos gestores e decisores terem uma perspectiva das áreas que devem controlar com total transparência e aumentar a sua capacidade de compreensão (do “porquê”);
- Suporte para a tomada de decisão – só uma compreensão oportuna da realidade pode permitir efetuar decisões eficazes; como tal, o conhecimento produzido pelos sistemas de BI, potenciados pelas tecnologias de comunicação actuais, deve suportar e justificar as medidas tomadas pelos vários intervenientes no processo de gestão.

Neste trabalho estudamos as principais soluções de suites open source BI que existem actualmente. Estas suites devem poder ser adoptadas pelas empresas, sem que essa implementação signifique um aumento dos custos, aumentando o conhecimento do seu negócio, que deverá ser traduzido num incremento da vantagem competitiva. O nosso principal objectivo é estudar as principais suites BI open source, gratuitas, existentes no mercado, analisá-las, comparando-as, apresentando as vantagens e desvantagens de cada uma. A análise será efectuada utilizando as últimas versões actualmente disponíveis das seguintes suites BI: JasperSoft [9], Palo [11], Pentaho [13], SpagoBI [16] e Vanilla [18].

Este artigo está estruturado da seguinte forma: na secção II é apresentada a análise das suites open

source Business Intelligence. Na secção III é explicada a metodologia utilizada, sendo descritos os critérios de avaliação e na secção IV é efectuada a comparação entre as suites analisadas. Por fim, na secção V são apresentadas as conclusões e proposto algum trabalho futuro.

II. Suites open source business intelligence

Suite é o termo usualmente utilizado para descrever ferramentas que agregam um conjunto de funcionalidades. Nesta secção estudaremos as suites que agregam as várias funcionalidades de BI, e que cumpram as necessidades para a implementação numa empresa. Apresentaremos a análise relativamente a cinco das principais suites BI: JasperSoft [9], Palo [11], Pentaho [13], SpagoBI [16] e Vanilla [18].

A. *JasperSoft*

A Jaspersoft afirma-se como fornecendo uma modalidade de self-service à medida das necessidades individuais das empresas [9]. A sua suite Jaspersoft foi criada em 2006, depois de vários anos a empresa ter criado as diversas ferramentas individualmente. A Jaspersoft afirma que fornece o conjunto de BI mais flexível, económico e amplamente implementado no mundo. No seu site, pode ler-se que já foi realizado o download de mais de 14 milhões de cópias do software de código aberto em todo o mundo, com 175.000 implementações de produção e mais de 14.000 clientes em 100 países. Afirmam ainda que a suite é regularmente actualizada por uma comunidade de desenvolvimento de mais de 225.000 membros registados. Possui a distribuição Community, open source, gratuita, e a distribuição Comercial distribuída por três edições (Express, Professional e Enterprise). A distribuição Community é bastante

limitada relativamente às distribuições comerciais e é distribuída segundo uma licença GNU GPL [10]. É composta pelos módulos JasperReports Server, JasperReports Library, JasperSoft ETL, JasperSoft Studio e iReport Designer.

O JasperReports Server é um servidor de relatórios stand-alone que pode ser incorporado em qualquer aplicação Java. Fornece relatórios e análises que podem ser incorporadas em aplicações Web ou móveis. Fornece informações em tempo real ou agendadas, para a Web, dispositivo móvel, impressora ou correio electrónico numa variedade de formatos. Está optimizado para partilhar, proteger e gerir de forma centralizada os relatórios e análises analíticas. Entre as funcionalidades, destacam-se: formatação e visualização de relatórios de forma interactiva, repositório centralizado e seguro, agendamento de geração e distribuição de relatórios e interface personalizável.

Segundo a JasperSoft, o JasperReports Library é ferramenta open source mais popular para criação de relatórios. É inteiramente programado em Java e é capaz de usar dados provenientes de qualquer tipo de origem e produzir documentos que podem ser visualizados, impressos ou exportados numa variedade de formatos, incluindo HTML, PDF, Excel, OpenOffice e Word.

A integração de dados (ETL – Extract, Transform, Load) é suportada pelo JasperSoft ETL. Este permite extrair dados de várias origens, transformar os dados com base em regras de negócio definidas e carrega-os numa Data Warehouse ou Data Mart para análise e extracção de relatórios. Entre as funcionalidades, destacam-se o ambiente de trabalho gráfico, mais de 500 ligações a componentes e o controlo de versões de trabalhos.

O JasperSoft Studio é um ambiente de desenho de relatórios baseado em Eclipse para JasperReports e JasperReports Server. Permite criar relatórios a partir de qualquer origem de dados, formatado para visualização no ecrã ou formato de impressão, com exportação para uma ampla gama de formatos. Entre as funcionalidades, destacam-se o ambiente de trabalho gráfico, os modelos de relatórios suportados por temas, integração com JasperReports Server, layouts sofisticados com gráficos, imagens, sub-relatórios e tabelas cruzadas, aceder aos dados através de JDBC, TableModels, JavaBeans, XML, Hibernate, CSV e origens personalizadas e publicar os relatórios em formato PDF, RTF, XML, XLS, CSV, HTML, XHTML, ficheiros de texto, DOCX ou OpenOffice.

Baseado em NetBeans, o iReport Designer, é uma ferramenta à imagem do JasperSoft Studio, basicamente com as mesmas funcionalidades. A versão mais recente do JasperSoft Studio é a 5.0, sendo a última actualização de Dezembro de 2012.

B. **Palo**

Palo é como é designada a suite open source desenvolvida desde 2007 pela empresa alemã Jedox AG [11]. Esta empresa desenvolve ainda uma versão comercial, com o nome Jedox, que tem normalmente um avanço no código desenvolvido entre 12 a 15 meses, relativamente à Palo. A suite Palo é composta por quatro aplicações de BI: Palo OLAP Server, Palo Web, Palo for Excel e ETL Server. A plataforma é completamente baseada em código open source, totalmente gratuita e distribuída sob uma licença GNU GPL (v2) [12].

As características do Palo, agrupam-se em três grupos: Excel Integration (Excel Add-in; OLAP); Web-Client (Web Spreadsheet; Web Pivot; File Manager; User Manager; ETL Manager); ETL

(ETL Server). É ainda composta pelos produtos Palo OLAP, Palo for Excel, Palo Web e Palo ETL.

O Palo OLAP é um servidor multidimensional OLAP, que trabalha em memória, estando os dados organizados em cubos, dimensões e elementos. Esta estrutura reflecte a forma como usualmente os utilizadores pensam o seu negócio. O facto de trabalhar em memória traduz-se numa vantagem relativamente a outros sistemas OLAP com suporte a uma base de dados relacional, que segundo os responsáveis chega a ser 100 vezes mais rápida, permitindo disponibilizar dados em tempo real.

O Palo for Excel é um add-in para o Excel e Open Office, que agregado ao OLAP Server Database constitui uma solução completa de BI. Permite integrar no Excel as dimensões relevantes da organização disponibilizadas pelo Palo, permitindo a sua visualização e análise ao mesmo tempo.

O Palo Web combina todos os componentes do Palo, que permite criar análises, planeamento ou relatórios, numa plataforma Web uniforme. Integra o Palo Spreadsheet, o Palo Pivot (realização de pivot tables ad-hoc), o Palo User Manager para gestão de acessos a utilizadores ou grupos, o Palo ETL Manager que cria e controla os projectos de integração de dados e, o Palo File Manager que disponibiliza a cada utilizador ou aplicação espaço para armazenamento de ficheiros, facilitando a integração e colaboração.

O Palo ETL permite realizar a integração de grandes quantidades de dados, através de ETL, a partir de origens heterogéneas. Baseado em ambiente Web, permite facilmente a conversão de um modelo relacional para um modelo OLAP, e realiza os *caches* necessários para o *drill-through* (ligação entre origens de dados), que exhibe

detalhadamente através de navegação. A versão mais recente do Palo é a 3.2, de Julho de 2011.

C. **Pentaho**

A suite Pentaho [13] é actualmente das mais usadas e das que maior reputação tem no mercado [14]. A Pentaho foi fundada em 2004 por um grupo de executivos com grande experiência em BI oriundos de empresas como a Business Objects, Cognos, Hiperion, JBoss, Oracle, Red Hat e SAS. Esta suite é composta pelas versões Enterprise e Community. A versão Enterprise é uma distribuição paga que contém todos os recursos, com interfaces amigáveis e de fácil compreensão para os utilizadores finais. A versão Community é open source, gratuita, e fornece praticamente os mesmos módulos, sendo que alguns não têm a interface tão amigável, e têm maior dificuldade de compreensão e implementação. O seu desenvolvimento é feito em Java, e pode ser executada a partir duma Java Virtual Machine. A versão Community é composta pelos módulos Pentaho Data Integration, Pentaho Reporting, Pentaho Dashboards, Pentaho Analysis Services e Pentaho Data Mining.

O Pentaho Data Integration, também conhecido como Kettle, permite realizar o ETL dos dados de forma fácil e intuitiva. Suporta uma grande biblioteca de objectos de mapeamento com suporte a diversas fontes de dados, e permite o armazenamento de dados para dimensões de Data Warehouses, outras bases de dados ou ficheiros. A transformação permite a limpeza de dados através de regras e a migração de dados entre aplicações.

O Pentaho Reporting é composto por duas ferramentas, uma de geração de relatórios, também conhecida como JFreeReport, e outra para geração de metadados, que permite a criação de relatórios ad-hoc via Web. Suporta várias fontes de dados, incluindo relacionais, OLAP e XML, permite

desenhar relatórios com suporte por *wizard*, multi-língua, filtros de visualização interactiva, e exportação para os formatos mais comuns.

Os indicadores de desempenho (KPI) são garantidos pelo Pentaho Dashboards, que permite a criação de painéis de controlo, e reunir na mesma janela os principais indicadores de um departamento ou de toda a empresa. As métricas são disponibilizadas de forma intuitiva, e permite integração com o Pentaho Reporting e Analysis. Disponibiliza ainda uma monitorização contínua com alertas de notificação de problemas.

O Pentaho Analysis Services, também conhecido como Mondrian, é um motor OLAP, baseado numa arquitectura ROLAP, que pode ser usado com os principais sistemas de gestão de base de dados, e possui funcionalidades como camada de metadados, linguagem MDX, cache em memória, tabelas agregadas, etc. Suportada num ambiente totalmente Web, permite a criação de relatórios com suporte drag&drop, visualização em gráficos, exportação da informação e visualização multi-dimensional com selecção de métricas e atributos a analisar.

Também conhecido como Weka, o Pentaho Data Mining é um conjunto de ferramentas para prospecção de dados, cujo conjunto de regras de classificação, regressão, associação e algoritmos de clustering permite ajudar a entender melhor o negócio e melhorar o desempenho futuro.

A versão Community é distribuída sob as licenças GNU General Public License versão 2.0 (GPLv2), GNU Lesser General Public License versão 2.0 (LGPLv2), e Mozilla Public License 1.1 (MPL 1.1) [15]. A versão mais recente desta suite é a 4.8, disponibilizada em Novembro de 2012.

D. **SpagoBI**

O SpagoBI é um software para desenvolvimento de projectos de BI em ambientes integrados. É suportado pela empresa italiana Engineering Group que, de acordo com o seu site, não tem interesse na sua exploração comercial, pois é uma empresa de consultadoria [16]. Para desenvolver a ferramenta a empresa criou a SpagoWorld Initiative em 2006, que mantém a ferramenta com uma comunidade open source, actualmente sob uma licença GNU Lesser General Public Licence (LGPL) [17]. Só existe uma versão do SpagoBI, gratuita e completa, ao contrário de outras suites que têm versões “community”, com funcionalidades reduzidas, e versões “enterprise”, com as funcionalidades necessárias a uma total integração numa organização. Toda a escalabilidade necessária numa organização é suportada pelo SpagoBI, a nível de arquitectura, funcionalidades e segurança.

A ferramenta permite uma visão de integração e não de produto, isto é, possibilita que, por exemplo, a funcionalidade OLAP seja usado pelo Mondrian ou o JPalo, ou que para gerar relatórios se use o Birt, Jasper ou Business Objects. Relativamente a segurança, o SpagoBI autoriza a criação de regras por grupos de utilizadores para a visualização de dados, e tem total integração LDAP e Active Directory.

Entre as principais funcionalidades, podem-se destacar a geração de relatórios sobre quatro motores distintos (JasperReport, Birt, Accessible Report, Business Objects), análise multidimensional (OLAP) suportada por três motores (JPivot/Mondrian, JPalo/Mondrian, JPXMLA), ou a geração de gráficos (JFreeChart, HChart, ExtChart) dos mais variados tipos (histogramas, barras, áreas, etc.). Quanto a indicadores de desempenho (KPI) permite a criação, gestão e exploração sobre

modelos hierarquizados, através de diferentes métodos, regras de cálculo, limites e regras de alarme. Relatórios ad-hoc (que possibilita que os utilizadores criem os seus próprios relatórios, com suporte a multi-página, tabelas simples e cruzadas, gráficos e diferentes tipos de layout), georeferenciação por GEO ou GIS (estabelecendo relação entre dados geográficos e dados da Data Warehouse), ou execução livre de pesquisas suportadas por interface Web são outras das funcionalidades disponibilizadas e que constituem uma mais-valia desta suite.

A SpagoBI é uma suite bastante completa que suporta ainda Data Mining (permitindo descobrir padrões de informação), ETL (Extract, Transform, Load) para carregar e gerir a Data Warehouse, Mobile BI (disponibilizando componentes específicos para interacção com equipamentos móveis, preparados para touch-screen). Destacam-se ainda a facilidade de trabalho colaborativo, integração com Open Office ou MS Office, e consolas de monitorização em tempo real, que podem ser integradas noutras aplicações ou processos.

Este conjunto de funcionalidades torna-a uma suite bastante completa e robusta, totalmente open source, cuja versão mais recente é a 3.6.0, disponibilizada em Novembro de 2012.

E. **Vanilla**

Vanilla é uma suite BI open source suportada pelo consórcio francês BPM-Conseil, com a primeira versão a ser lançada em Junho de 2008 [18]. Fundada em 2004, a BPM-Conseil possui vasta experiência em consultadoria, e com a sua equipa de programadores iniciou o desenvolvimento do Vanilla em 2007. Mais que uma agregação de ferramentas, o Vanilla é uma plataforma que permite definir a sequência completa do negócio

graças à grande diversidade de componentes que possui. Sendo uma suite disponibilizada integralmente em open source, é distribuída sob uma licença MPL (Mozilla Public Licence) [19], e desenvolvida em linguagem Java.

A suite faculta acesso a suporte técnico e a um fórum, mantidos pela equipa de desenvolvimento. Suporta Unicode para a base de dados de cliente e um repositório com quatro línguas padrão: Inglês, Francês, Espanhol e Português. É composta por uma grande diversidade de componentes de onde se destacam um Web portal, que é o servidor aplicativo Tomcat Vanilla e possui o Vanilla Web Portal, o motor runtime para objectos BI integrado no portal Vanilla, os módulos Web para utilizadores finais e também alguns módulos para Web designers. FreeMetadata, Vanilla plugins para Birt, Free Dashboards, FreeWebReport, Analysis Schema Designer, BiGateway para ETL, BiWorkflow e BiProcessManager são as outras componentes da suite.

Entre as suas principais características podemos destacar a integração segura de metadados com o Birt com optimizações SQL, bem como com o FreeDashboards e Cubes. Permite ainda integração com ETL e Workflow. Possui uma base de dados de indicadores de desempenho (KPI), totalmente integráveis com outros módulos como o BiGateway ou Dashboard. Possibilita ainda realizar todas as etapas padrão de ETL, com suporte a SQL, OLAP, FreeMetrics e origens de dados FreeMetada.

O suporte para SaaS (Software as a Service) permite hospedar diferentes projectos num único servidor, disponibilizando ainda multi-repositórios. A arquitectura de cluster é suportada pelo módulo Hypervision, e permite gerir clusters de diferentes serviços Vanilla (ETL, relatórios, etc.).

A suite é bastante completa e robusta, com a vantagem de ser totalmente open source, e entre as diversas funcionalidades que disponibiliza ainda se destaca o suporte a Workflow BPMN 2.0 integrado para desenho de processos de negócio complexos, bem como inúmeros recursos de BI como colunas dinâmicas, origens alternativas de dados, modos de ruptura para validar políticas de segurança e cálculos personalizáveis no FreeMetadata.

Suporta ainda a gestão integrada de documentos, e permite a integração completa de módulos e documentos: Metadata, Dashboard, Report, Cubes, KPI, Workflow e ETL. A versão mais recente do Vanilla é a 4.2, lançada em Janeiro de 2013.

III. Metodologia de avaliação

A avaliação das ferramentas foi baseada nos critérios que a Gartner definiu no seu Magic Quadrant de 2013 [14]. Estes critérios foram por nós adaptados para a realidade das suites open source, e traduzem-se em 11 capacidades divididas por 3 categorias de funcionalidades: i) Integração, ii) Disponibilidade (entrega) da informação e iii) Análise. Os critérios de classificação para cada uma das capacidades, são:

A. Integração

- ETL – As ferramentas de ETL permitem efectuar a extracção de dados de diversos sistemas, transformar esses dados segundo as regras de negócio, e efectuar o seu carregamento para Data Marts ou Data Warehouses.
- Colaboração – Esta capacidade permite que os utilizadores de BI partilhem e discutam a informação disponibilizada, com *chat* e notas, e que a possam integrar noutras ferramentas colaborativas ou sociais.

B. Disponibilidade da Informação

- Reporting – Disponibiliza a capacidade de criar reports interactivos formatados, com ou sem parâmetros, com distribuição altamente escalável e capacidades de agendamento. Adicionalmente deve disponibilizar um

largo conjunto de estilos de reports pré-definidos (por exemplo, financeiros, dashboards operacionais e de desempenho), e deve permitir aos utilizadores o acesso e interacção com os conteúdos disponibilizados em diversas plataformas incluindo Web, dispositivos móveis, e ambientes comuns de portal.

- Dashboards – Este subconjunto de reporting inclui a capacidade de publicar formalmente, baseado em Web ou para dispositivos móveis, reports com mostradores interactivos de informação. Estes mostradores indicam o estado das métricas de desempenho comparadas com um objectivo. São ainda usados para disseminar dados em tempo real de aplicações operacionais ou em conjunto com complexos eventos de processamento.
- Pesquisas ad-hoc – Esta capacidade permite que os utilizadores façam as suas perguntas de forma livre, sem solicitar a criação de um report. As ferramentas devem possuir uma camada de semântica robusta para permitir que os utilizadores naveguem pelas origens de dados.
- Integração com Office – Em muitos casos as plataformas de BI são usadas como uma camada intermédia para gerir, proteger e executar tarefas de BI, sendo o Office (particularmente o Excel) a agir como cliente de BI. Nestes casos, é vital que as ferramentas permitam integração com as aplicações do Office, incluindo suporte para formatos de documentos e apresentações, fórmulas, refreshamento de dados e tabelas dinâmicas. As características “cell locking” e “write-back” serão benefícios adicionais e tidas como integração avançada.
- Mobilidade BI – Esta capacidade permite que as organizações apresentem reports e dashboards nas plataformas móveis, numa publicação e/ou modo interactivo, aproveitando o modo de interacção dos dispositivos e outras capacidades não disponíveis normalmente nos monitores.

C. **Análise**

- OLAP – Permite que os utilizadores finais analisem dados com pesquisas bastante rápidas e desempenho de cálculo, navegando facilmente nos caminhos de pesquisa multidimensionais. Nalguns casos permite o write-back de valores para bases de dados proprietárias para planeamento e modelação. Pode abranger uma variedade de arquitecturas de dados (relacional, multidimensional) e de armazenamento (disco ou memória).
- Visualização interactiva – Oferece ao utilizador a possibilidade de visualizar numerosos aspectos dos dados mais eficientemente, usando imagens e gráficos, em vez de linhas e colunas. Ao longo do tempo, a visualização avançada permitirá fatiar os dados para os incluir mais nos processos orientados a projectos de BI, permitindo que todos os intervenientes

compreendam melhor o fluxo de trabalho através da representação visual.

- Modelo preditivo e data mining – Permite explorar grandes quantidades de dados à procura de padrões consistentes, como regras de associação ou sequências temporais, para detectar relacionamentos entre variáveis, detectando assim novos subconjuntos de dados, usando técnicas matemáticas avançadas. Os programadores de projectos BI podem integrar os modelos facilmente em reports, painéis de análise e processos de negócio.
- Scorecards - Torna as métricas exibidas em painéis de análise em algo mais avançado, aplicando-as num mapa estratégico que alinha os indicadores de desempenho (KPI) com objectivos estratégicos. As métricas dos scorecards podem ser inseridas em reports e informações relacionadas, a fim de realizar análises mais aprofundadas. Implica o uso de metodologias de gestão de desempenho.

IV. **Comparação das suites BI**

Após o estudo das suites open source BI que identificámos como tendo capacidades nas três categorias de funcionalidades estudadas, foi efectuada uma comparação relativamente a estas funcionalidades. De referir que há alguns anos, Thomsen & Pedersen também fizeram uma análise de várias ferramentas open source para BI, contudo não efectuaram nenhuma comparação entre as ferramentas [20]. Também Golfarelli [21] realizou um estudo semelhante, com menos suites em análise, comparando-as entre a versão comercial e open source. Nesse estudo foram evidentes as diferenças de funcionalidades entre as duas versões, não sendo as open source tão sofisticadas. Na Tabela 1 é apresentada a comparação das suites tendo em conta as capacidades suportadas.

TABELA 1 COMPARAÇÃO DAS SUITES OPEN SOURCE BI

Capacidades	Suites Open Source BI				
	<i>JasperSoft</i>	<i>Palo</i>	<i>Pentaho</i>	<i>SpagoBI</i>	<i>Vanilla</i>
ETL	✓	✓	✓	✓	✓
Colaboração		✓		✓	✓

Reporting	✓	✓	✓	✓	✓
Dashboards	✓	✓	✓	✓	✓
Pesquisas ad-hoc		✓	✓	✓	✓
Integração com Office	✓	✓	✓	✓	
Mobilidade BI	✓		✓	✓	✓
OLAP	✓	✓	✓	✓	✓
Visualização interactiva	✓	✓	✓	✓	✓
Data Mining			✓	✓	✓
Scorecards			✓	✓	✓

Desta análise, podemos destacar a SpagoBI como sendo a ferramenta mais completa, que suporta todas as capacidades por nós seleccionadas, sendo que das suites estudadas, é, a par com a Vanilla, uma suite que possui uma única versão, totalmente open source e gratuita. A SpagoBI disponibiliza ainda outras capacidades além das avaliadas neste estudo, como por exemplo localização inteligente (GEO-GIS) ou Master Data Management (para usufruto das capacidades write-back na base de dados) e rivaliza mesmo com a maior parte das versões comerciais das outras suites em estudo.

A Pentaho surge como outra ferramenta bastante poderosa na nossa análise, visto que suporta basicamente as mesmas capacidades da SpagoBI. Contudo, possui uma versão comercial, o que consideramos uma desvantagem. É no entanto, das estudadas, a suite eventualmente mais usada, como referido pela Gartner [11]. Oferece, junto com a SpagoBI uma solução completa que permite atingir os objectivos duma implementação BI numa empresa, em busca dum maior conhecimento do negócio e duma vantagem competitiva. Contudo, do estudo efectuado concluímos que a suite não permite que os utilizadores partilhem e discutam a informação disponibilizada, com *chat* e notas, e que a possam integrar noutras ferramentas colaborativas.

O facto de não suportar uma integração com Office (Microsoft, OpenOffice, LibreOffice ou outros) traduz-se numa desvantagem para a Vanilla, que,

sendo das suites menos conhecidas, mas que nas restantes funcionalidades rivaliza tanto com a SpagoBI como com a Pentaho. É junto com estas uma boa solução de implementação de BI numa empresa.

Já a Palo e a JasperSoft, na sua versão open source, apresentam um número mais restrito de funcionalidades. O facto de não suportarem Data Mining nem Scorecards deixa-as em grande desvantagem, pois negligenciam nas versões open source capacidades fundamentais nas implementações de BI, apostando essencialmente nas versões comerciais, nas quais disponibilizam outro tipo de capacidades. O facto da Palo não disponibilizar uma versão desde Julho de 2011 coloca-a, no nosso entender, para a última posição deste estudo, pois não disponibiliza soluções de mobilidade, tão usadas actualmente, e que são uma tendência do mercado a par com a utilização de Big Data (não levado em conta neste estudo). A JasperSoft peca essencialmente pela discrepância de oferta da versão open source relativamente às suas versões comerciais, que são três, apostando visivelmente mais nessa vertente. No entanto, relativamente às capacidades disponibilizadas na versão Community, estas são robustas e competitivas com as restantes.

v. Conclusões e trabalho futuro

Neste estudo efectuámos a análise das últimas versões das principais suites open source BI existentes no mercado. Estas suites tinham que satisfazer um conjunto de capacidades que definimos como fundamentais para uma implementação BI numa empresa sem que os custos aumentassem, e analisámo-las baseadas nesses critérios. O estudo foi realizado tendo por base a

informação disponibilizada pelas próprias suites nos seus sites e em documentação técnica sobre as mesmas.

A nossa principal conclusão é que as suites open source BI têm já um alto nível de confiabilidade e devem ser consideradas uma alternativa válida em relação às versões proprietárias. Isto é ainda mais evidente para as organizações em que a quantidade de dados e o volume de trabalho não são pontos críticos. Da nossa análise, concluímos que as suites Open Source Business Intelligence estão crescendo tanto em funcionalidades, como qualidade e aspecto visual. Das cinco suites analisadas, a Palo é sem dúvida a menos completa, e a SpagoBI e a Pentaho são aqueles com maior potencial para utilização numa empresa. Destas duas destacamos a Spago BI porque além das suas capacidades é aquela que apenas disponibiliza uma única versão totalmente open source e gratuita o que garante uma aposta sólida neste segmento.

Como trabalho futuro pretendemos continuar este estudo avaliando as suites noutros aspectos técnicos e de usabilidade. Pretendemos ainda realizar a implementação em ambiente real, numa empresa, avaliando o desempenho das várias suites nessas condições.

Referências

- [1] Tzu, Sun, "A Arte da Guerra", Texto Editores, Julho 2011
- [2] M. Porter, V. Millar, "How information gives you competitive advantage", Harvard Business Review, July-August 1985
- [3] S. Williams, N. Williams, "The Business Value of Business Intelligence", Business Intelligence Journal, Fall 2003
- [4] D. J. Power, "A Brief History of Decision Support Systems, version 4.0," DSSResources.COM. World WideWeb, <http://dssresources.com/history/dsshhistory.html>, version 4.0, March 10, 2007.
- [5] H. Luhn, "A Business Intelligence System", IBM Journal, October, 1958.
- [6] L. Amaral, J. Varajão, "Planeamento de Sistemas de Informação - 4ª Edição Actualizada e Aumentada", FCA – Editora de Informática, 2007, pp. 6-9.
- [7] R. Antonelli, "Conhecendo o Business Intelligence", CAP Accounting and Management, Revista Científica do TECAP, N°3, Ano 3, Vol. 3, 2009, pp. 79-85.
- [8] C. Sezões, J. Oliveira, M. Baptista, "Business Intelligence", SPI – Sociedade Portuguesa de Inovação, 2006. Disponível em: http://www.bi4all.pt/docs/Manual_Business_Intelligence.pdf, 20-02-2013.
- [9] <http://www.jaspersoft.com/>, acedido em 20-02-2013
- [10] <http://community.jaspersoft.com/>, acedido em 20-02-2013
- [11] <http://www.palo.net/>, acedido em 20-02-2013
- [12] <http://sourceforge.net/projects/palo/>, acedido em 20-02-2013
- [13] <http://www.pentaho.com/>, acedido em 20-02-2013
- [14] K. Schlegel, R. L. Sallam, D. Yuen, and J. Tapadinhas, "Magic Quadrant for Business Intelligence and Analytics Platforms", Gartner, February 2013.
- [15] <http://en.wikipedia.org/wiki/Pentaho>, 20-02-2013
- [16] <http://www.spagoworld.org/xwiki/bin/view/SpagoBI/Vision>, acedido em 20-02-2013.
- [17] <http://www.spagoworld.org/xwiki/bin/view/SpagoBI/Licenses>, acedido em 20-02-2013.
- [18] <http://www.bpm-conseil.com/>, acedido em 20-02-2013.
- [19] http://fr.wikipedia.org/wiki/Vanilla_:Plateforme_d%C3%A9cisionnelle, acedido em 20-02-2013.
- [20] C. Thomsen, T. Pedersen (2009), "A survey of open source tools for business intelligence", International Journal of Data Warehousing and Mining (IJDWM), 5(3), 56-75.
- [21] M. Golfarelli, "Open Source BI Platforms: a Functional and Architectural Comparison", DaWaK, pp. 287-297, Springer, 2009

Anexo D

Artigo “OpenBRR Evaluation of an Open Source BI Suite”

C3S2E 2013 - Sixth International C* Conference on Computer Science & Software Engineering
10-12 de Julho de 2013, Porto - Portugal

OpenBRR Evaluation of an Open Source BI Suite

António Marinheiro
Polytechnic Institute of Coimbra
ISEC - Coimbra Institute of Engineering
Rua Pedro Nunes, 3030-199 Coimbra, Portugal
Tel. ++351 239 790 200

a21200015@alunos.isec.pt

Jorge Bernardino
Polytechnic Institute of Coimbra
ISEC - Coimbra Institute of Engineering
Rua Pedro Nunes, 3030-199 Coimbra, Portugal
Tel. ++351 239 790 200

jorge@isec.pt

ABSTRACT

Business Intelligence (BI) refers to technologies that target how business information is collected analyzed and presented. Combining these features results in software called Business Intelligence Suites. These systems help companies and organizations to know their business, providing better decision support, which reflects into a competitive advantage. Open source suites have the advantage of not increasing costs, but raise some doubts about its adoption. Evaluate open source software under a recognized method is important to ensure its quality. In this paper we evaluate an Open Source Business Intelligence Suite, using OpenBRR (Business Readiness Rating for Open Source), an open source software assessment methodology.

Categories and Subject Descriptors

H.4 [Information Systems Applications]. H.4.2 [Types of Systems]. H.4.2 [Decision Support].

General Terms

Management, Measurement, Experimentation, Standardization.

Keywords

Open Source Business Intelligence Suites; Open Source Software Assessment Methodologies; OpenBRR.

1. INTRODUCTION

The term Business Intelligence (BI) first appeared in 1958, defined by H. P. Luhn in an article on IBM Journal [1]. This term was taken again and redefined by Howard Dresner

from Gartner Group in 1989 as being a set of concepts and methods to improve enterprising decision taken by companies turning to supporting systems based on facts [2]. In other words this set up of concepts and methods that constitute the BI support and increase the capacity of decision taken by the organizations transforming their data in knowledge [3]. This capacity is only possible with the support of BI tools, either self-sufficient or gathered into an integrated platform like BI suites. BI suites should be able to be adopted by the enterprises without meaning an increasing of their costs, otherwise promoting their business and enlarging the knowledge of their affairs translated in a favorable of their competitiveness. Pentaho Community Edition is one of the Open Source BI suites with a broader use worldwide [4]. However, the fact of being open source raises some doubts. Is the software suitable for its use? How to assess the degree of maturity of the software?

In order to assuring its quality over last years emerged several methodologies of assessment of open source software which should be taken in consideration by the organizations before

considering the adoption of one of these software. Two of these methodologies that stand out are Qualification and Selection Open Source (QSOS) [5] and Business Readiness Rating (OpenBRR) [6]. Both assessment methodologies help its users to reach a similar target, which is selecting from similar projects the most suitable to their context. Deprez [7] points out on the advantages and disadvantages of each methodologies stating that OpenBRR enables tailoring the criteria to adapt them to a context and that QSOS is ambiguous in more than half of its criteria. This analysis has taken us to decide for the adoption of the Open BRR methodology to assess Pentaho Open Source BI Suite.

The remainder of this paper is structured as follows. In section 2, we explain the fundamentals of open BRR methodology. Section 3 presents the Pentaho assessment using OpenBRR. Finally, in section 4 we present our conclusions and point out future work.

2. OPEN BRR

Open Business Readiness Rating (OpenBRR) is an open source software assessment methodology defining an open and standard assessment process. OpenBRR methodology is based on the assessment over 12 categories: Functionality, Usability, Quality, Security, Performance, Scalability, Architecture, Support, Documentation, Adoption, Community and Professionalism, having each category several metrics to be assessed.

In this level the grade ranking may have different levels of importance depending on the requirements to use the software and the

assessment is split in 4 phases: 1) Identifying software components list to be analyzed, estimating each component in relation to assessment criteria and removing from analyzing any software component that does not satisfy the use requirements; 2) assigning weights to categories and metrics: a) Rank the 12 categories in according to importance (1- the highest, 12 – lowest); b) take the top 7 (or fewer) categories and assign a percentage of importance to each one totaling 100%; c) For each metric within a category rank the metric in accordance to its importance; d) To each metric within a category assign the importance by percentage, totaling all the metrics 100% of the category; 3) gather data for each metric used in each category and calculate its weighting in a range between 1 (unacceptable) to 5 (excellent); 4) Finally, on the fourth stage, the qualification of the category and the weighting factors should be used to calculate the BRR final score.

The category of functionality is assessed differently from the others and follows an own method. In this category we pretend to analyze and evaluate the characteristics which the tools have, or should have (what is being considered as expectable), more extras. On the case of open source BI suites the ETL, reporting, etc. The method to assess this category is as follows: 1) Set down the resources (characteristics) to analyze, valuating them from 1 to 3 (less important to very important); 2) Classify them in a cumulative sum (from 1 to 3) enabling us to give a negative grade (withdraw points to the cumulative sum) if they do not meet a characteristic; 3) Do the same to extra resources; 4) Divide the cumulative sum (including extras)

by the maximum score obtainable by the base characteristics (without extras). This intentionally may be lower to 0 (zero) or higher than 100, valuing the extras and devaluing (punishing) the basic not considered characteristics; 5) Standardize the prior result to a scale from 1 to 5 to the Functionality category on the following scale: a) under 65% score=1 (unacceptable); b) 65% - 80% score=2 (poor); c) 80% - 90% score=3 (acceptable); d) 90% - 96%, score=4 (good); e) over 96%, score=5 (excellent). This final scale fits then in the same measurement of the remaining characteristics.

3. OPENBRR PENTAHO EVALUATION

Having in mind the analyses interest the most relevant 7 categories and its weight (importance) were: Functionality (40%), Usability (15%), Quality (15%), Documentation (10%), Community (10%), Safety (5%) and Architecture (5%).

To assess the category of Functionality the more common capacities of BI suites were considered, assigning its relevance on the scale from 1 (irrelevant) to 3 (very important). Table 1 shows the result of the final assessment of category Functionality.

Table 1. Rating of functionality category

Capabilities	Weight	Rating
ETL	3	3
Collaboration	2	2
Reporting	3	3
Dashboards	3	3
Ad-hoc query	2	2
Office integration	2	0
Mobile BI	2	2

OLAP	3	3
Interactive	1	1
Data Mining	1	1
Scorecards	2	2
Extra		
Geo tagging	2	0
Total	26	22

On assessing of Functionalities according the Open BRR scale, Pentaho will give mark 3 corresponding to 84.6%.

For the remaining categories the criterion established by the method was used, with assessment shown on Table 2.

Table 2. Rating the other categories

Categories	Weight	Rating
Usability	15%	5
Quality	15%	1,9
Documentation	10%	5
Community	10%	3,8
Security	5%	5
Architecture	5%	5

Pondering each category marks we get the final result that is shown in Table 3.

Table 3. OpenBRR final score

Categories	Weight	Weighting
Functionality	40%	1,2
Usability	15%	0,75
Quality	15%	0,285
Documentation	10%	0,5
Community	10%	0,38
Security	5%	0,25
Architecture	5%	0,25
Total		3,615

Using OpenBRR method, choosing the more relevant categories and having in account the

intended implementation and sharing its weights, Pentaho Community Edition reaches the mark 3,615, which is considered “good” software.

4. CONCLUSIONS AND FUTURE WORK

The application of OpenBRR methodology becomes useful in this assessment as it offers an objective evaluation criterion, reproducible and auditable, assigning a numeric value allowing in this way the comparison with other evaluations of other products. The evaluation of Pentaho Community Edition shows good marks and in the final score it is rated as “good” software. As future work, we intend to evaluate more open source business intelligence suites in order to set

up a broader and measurable knowledge whose adoption is useful for enterprises.

5. REFERENCES

- [1] Luhn, H. 1958. A Business Intelligence System, IBM Journal (Oct. 1958).
- [2] Power, D. J. 2007. *A Brief History of Decision Support Systems, version 4.0.* (Mar. 2007). <http://dssresources.com/history/dsshhistory.html>.
- [3] Sezões, C., Oliveira, Baptista, J., M. 2006. Business Intelligence. *SPI – Sociedade Portuguesa de Inovação*.
- [4] <http://community.pentaho.com/>
- [5] <http://www.qsos.org/>
- [6] “Business Readiness Rating for Open Source - A Proposed Open Standard to Facilitate Assessment and Adoption of Open Source Software”. http://master.librosoft.es/sites/default/files/Materiales_MSWL_2010_2011/Project%20Evaluation/materiales/OpenBRR_Whitepaper.pdf
- [7] Deprez, J. C., Alexandre, S. 2008. Comparing Assessment Methodologies for Free/Open Source Software: OpenBRR & QSOS. In *CETIC* (Charleroi, Belgium, 2008).