



Instituto Politécnico de Coimbra  
Instituto Superior de Contabilidade  
e Administração de Coimbra

Nuno Rafael Almeida Leite

*Business Intelligence no Suporte à Decisão:  
Soluções Open Source*

*Business Intelligence no Suporte à Decisão: Soluções Open Source*

Nuno Rafael Almeida Leite

ISCAC | 2018

Coimbra, setembro de 2018





Instituto Politécnico de Coimbra  
Instituto Superior de Contabilidade  
e Administração de Coimbra

Nuno Rafael Almeida Leite

*Business Intelligence no Suporte à Decisão:  
Soluções Open Source*

Trabalho de Projeto submetido ao Instituto Superior de Contabilidade e Administração de Coimbra para cumprimento dos requisitos necessários à obtenção do grau de **Mestre em Sistemas de Informação de Gestão**, realizado sob a orientação do Professor Jorge Fernandes Rodrigues Bernardino e coorientação da Professora Isabel Maria Mendes Pedrosa.

Coimbra, setembro de 2018

## **TERMO DE RESPONSABILIDADE**

Declaro ser o autor deste projeto, que constitui um trabalho original e inédito, que nunca foi submetido a outra Instituição de ensino superior para obtenção de um grau acadêmico ou outra habilitação. Atesto ainda que todas as citações estão devidamente identificadas e que tenho consciência de que o plágio constitui uma grave falta de ética, que poderá resultar na anulação do presente projeto.

## AGRADECIMENTOS

Com o terminar deste trabalho, concluo uma etapa importante de evolução e consolidação de conhecimento numa área nova para mim: Sistemas de Informação, em particular *Business Intelligence*.

Do início à conclusão desta etapa importa agradecer:

Ao meu orientador Professor Jorge Bernardino e coorientadora Professora Isabel Pedrosa pelo contínuo acompanhamento, pelo espírito crítico, pelas sugestões, pela motivação e pelo apoio em momentos menos fáceis na elaboração deste trabalho.

À minha família por todo apoio incondicional.

Ao João Carriço, Miguel Cruz, Bruno Duarte e Gabriel Rosa por me terem proporcionado a oportunidade de aceder ao mundo do *Business Intelligence* e pelo apoio no desenvolvimento das minhas ideias.

Aos meus professores e colegas de Mestrado que foram uma parte importante neste caminho.

Por fim, mas não menos importante, à Cláudia Moniz pelo incentivo à realização deste Mestrado e por todo o apoio em todas as fases deste caminho.

## RESUMO

O conhecimento que se consegue extrair dos dados existentes nas organizações é fundamental para aumentar a sua vantagem competitiva nos atuais mercados em constante mutação. Esta constante mudança leva as organizações a tomar decisões estratégicas, táticas ou operacionais com maior frequência. Há, portanto, uma necessidade crescente de agilidade neste processo e que as decisões sejam menos intuitivas e mais fundamentadas em informação credível. É neste contexto que os sistemas de *Business Intelligence* (BI) surgem com o objetivo de auxiliar neste processo de tomada de decisão, recolhendo dados, processando-os/analizando-os e disponibilizando a informação e conhecimento que deles se retira. Se no início existia um domínio por parte das aplicações comerciais, que, por si só, representava um custo normalmente elevado apenas com licenciamento, tal já não é completamente verdade. As soluções *open source* têm ganho expressão de uma forma geral e na área do *Business Intelligence* não é diferente, sendo atualmente uma alternativa viável às soluções comerciais. Com o objetivo de promover esta alternativa, neste trabalho identificámos as principais plataformas *open source* de BI: Jaspersoft, Knowage e Pentaho. Foram posteriormente submetidas a um processo de avaliação para identificar a que atualmente se apresenta como melhor opção. Para tal foi utilizada a metodologia de avaliação de software *open source* OSSpal. Neste processo de avaliação identificámos a Pentaho como melhor opção atual dentro das plataformas *open source* de BI disponíveis. Posteriormente, realizámos um estudo de caso aplicado a uma empresa da área logística, com a utilização prática da Pentaho de forma a comprovar a viabilidade destas soluções *open source*. Concluímos que a Pentaho deu resposta integral aos problemas propostos, ainda que, em parte, se tenha recorrido a um *plugin* desenvolvido pela comunidade de utilizadores e que não está disponível de base na plataforma. A vasta comunidade de utilizadores, fruto da adoção mundial da Pentaho, é um ponto forte nesta plataforma *open source*, pois potencia o aparecimento e a evolução de *plugins*, bem como constitui um importante grupo de suporte a utilizadores menos experientes. Concluímos assim que as organizações devem equacionar a Pentaho num momento de escolha de uma plataforma *open source* de *Business Intelligence*.

Palavras-chave: *Business Intelligence*; *Open Source*; OSSpal

## **ABSTRACT**

The knowledge one can extract from existing data in organizations is critical to increasing their competitive advantage in today's changing markets. This constant change leads organizations to make strategic, tactical, or operational decisions more often. There is therefore a growing need for agility in this process and for decisions to be less intuitive and more based on credible information. It is in this context that Business Intelligence (BI) systems arise with the objective of assisting in this decision-making process, collecting data, processing / analysing them and making available the information and knowledge withdrawn from them. If at first there was a dominance by commercial applications, which in itself represented a normally high cost only with licensing, this is no longer completely true. Open source solutions have gained expression in a general way and in the area of Business Intelligence is no different and is currently a viable alternative to commercial solutions. With the objective of promoting this alternative, we have identified the main open source BI platforms: Jaspersoft, Knowage and Pentaho. They were subsequently submitted to an evaluation process to identify what is currently the best option. For this, the OSSpal open source software evaluation methodology was used. In this evaluation process we have identified Pentaho as the best current option within the open source BI platforms available. Subsequently, we conducted a case study applied to a logistics company, with the practical use of Pentaho in order to prove the viability of these open source solutions. We conclude that Pentaho has given an integral answer to the problems proposed, although, in part, it has resorted to a plugin developed by the community of users and that is not originally available in the platform. The vast community of users, as a result of the worldwide adoption of Pentaho, is a strong point in this open source platform, as it enhances the appearance and evolution of plugins, as well as being an important support group for less experienced users. We conclude that organizations should consider Pentaho when choosing an open source Business Intelligence platform.

Keywords: Business Intelligence; Open Source; OSSpal

# ÍNDICE GERAL

1	INTRODUÇÃO .....	1
1.1	Objetivos deste trabalho .....	2
1.2	Metodologia .....	3
1.3	Contribuições deste trabalho .....	3
1.4	Estrutura do trabalho .....	4
2	<i>BUSINESS INTELLIGENCE</i> .....	7
2.1	Importância do <i>Business Intelligence</i> na tomada de decisão .....	9
2.2	Arquitetura de sistemas de <i>Business Intelligence</i> .....	11
2.3	Conceitos associados a um Sistema de <i>Business Intelligence</i> .....	12
2.3.1	Sistema de Dados Operacionais.....	13
2.3.2	Processos ETL .....	13
2.3.3	<i>Data Warehouse</i> .....	14
2.3.4	Modelo Multidimensional.....	16
2.3.5	Sistemas OLAP ( <i>On-line Analytical Processing</i> ).....	17
2.3.6	<i>Data Mining</i> .....	20
2.3.7	<i>Dashboards</i> .....	21
3	TRABALHOS RELACIONADOS .....	23
3.1	Plataformas <i>Open Source de Business Intelligence</i> .....	23
3.2	Metodologias de Avaliação de <i>Software Open Source</i> .....	29
4	PLATAFORMAS <i>OPEN SOURCE DE BUSINESS INTELLIGENCE</i> .....	31
4.1	Jaspersoft.....	31
4.2	Knowage (SpagoBI).....	33
4.3	Pentaho.....	36



5	METODOLOGIAS DE AVALIAÇÃO DE <i>SOFTWARE OPEN SOURCE</i> .....	39
5.1	Escolha de Metodologia.....	39
5.2	OSSpal.....	40
6	SELEÇÃO DA PLATAFORMA <i>OPEN SOURCE DE BI</i> .....	43
6.1	Comparação das plataformas .....	43
6.2	Avaliação das plataformas .....	46
7	ESTUDO DE CASO.....	51
7.1	Preservação de anonimato.....	51
7.2	A empresa.....	51
7.3	Contexto Operacional.....	51
7.4	Análise de requisitos .....	52
7.5	Tratamento de dados .....	54
7.5.1	Dados fornecidos pela Good Mailman .....	54
7.5.2	Campos adicionais .....	54
7.5.3	Modelo Multidimensional.....	55
7.5.4	Dimensão Data Calendário .....	57
7.6	Instalação da plataforma Pentaho.....	57
7.7	Implementação do Modelo Multidimensional e cubo OLAP .....	58
7.8	Criação de Análises.....	59
7.8.1	Coordenador de Centro.....	59
7.8.2	Coordenação Geral .....	62
7.8.3	<i>Dashboard</i> .....	64
7.9	Conclusões .....	65
8	CONCLUSÕES E TRABALHO FUTURO .....	67
	REFERÊNCIAS BIBLIOGRÁFICAS .....	71
	ANEXOS .....	79

ANEXO 1 Avaliação Comparativa de Plataformas de <i>Business Intelligence Open Source</i> para PME – CISTI 2018 .....	81
ANEXO 2 Open Source Business Intelligence Platforms’ Assessment using OSSpal Methodology – ICE-B 2018 .....	89
ANEXO 3 Preparação da Plataforma Pentaho .....	97
ANEXO 4 Código SQL das Tabelas criadas em PostgreSQL.....	103
ANEXO 5 Código SQL das queries utilizadas no PRD .....	109
ANEXO 6 Código MDX das queries utilizadas no CDE .....	113

## ÍNDICE DE FIGURAS

<b>Figura 2.1</b> Papel do BI na tomada de decisão <b>Fonte:</b> Olszak & Ziembra (2007, p.137)	10
<b>Figura 2.2</b> Arquitetura geral de um sistema de Business Intelligence <b>Fonte:</b> Adaptado de Chaudhuri et al. (2011)	12
<b>Figura 2.3</b> Exemplo de Esquema em Estrela <b>Fonte:</b> Santos & Ramos (2009)	16
<b>Figura 2.4</b> Exemplo de Esquema em Constelação <b>Fonte:</b> Santos & Ramos (2009)	17
<b>Figura 2.5</b> Exemplo de Esquema em Floco de Neve <b>Fonte:</b> Santos & Ramos (2009)	17
<b>Figura 2.6</b> Exemplo de um cubo de 3 dimensões <b>Fonte:</b> Kamber et al. (2011)	18
<b>Figura 2.7</b> Manipulação de cubos <b>Fonte:</b> Kamber et al. (2011)	19
<b>Figura 4.1</b> Arquitetura da Plataforma Jaspersoft <b>Fonte:</b> (Jaspersoft, 2018a)	32
<b>Figura 4.2</b> Interface JaspersoftReports Server <b>Fonte:</b> (Jaspersoft, 2018b)	33
<b>Figura 4.3</b> Exemplo de relatório da plataforma Knowage <b>Fonte:</b> (Knowage, 2018b)	34
<b>Figura 4.4</b> Arquitetura da plataforma Knowage (SpagoBI) <b>Fonte:</b> (SpagoBI, 2014)	35
<b>Figura 4.5</b> Exemplo de relatório da plataforma Pentaho <b>Fonte:</b> (Pentaho, 2018a)	36
<b>Figura 4.6</b> Arquitetura da plataforma Pentaho <b>Fonte:</b> (Pentaho, 2018a)	37
<b>Figura 5.1</b> Metodologia OSSpal <b>Fonte:</b> Adaptado de Wasserman (2014)	41
<b>Figura 7.1</b> Excerto dos registos fornecidos pela Good Mailman	54
<b>Figura 7.2</b> Excerto da Tabela da Dimensão Cliente	56
<b>Figura 7.3</b> Modelo Multidimensional em estrela	56
<b>Figura 7.4</b> Excerto da Tabela da Dimensão Data	57
<b>Figura 7.5</b> Tabelas criadas em PostgreSQL	58
<b>Figura 7.6</b> Cubo contruído com o PSW	59
<b>Figura 7.7</b> Layout do relatório desenvolvido no PRD	60
<b>Figura 7.8</b> Exemplo de relatório por cliente	60
<b>Figura 7.9</b> Interface de passagem de parâmetros ao relatório	61

<b>Figura 7.10</b> Exemplo de relatório por código postal .....	61
<b>Figura 7.11</b> Interface do Saiku com ligação ao cubo Good Mailman ativa .....	62
<b>Figura 7.12</b> Análise da evolução semanal por cliente obtida com recurso ao Saiku.....	63
<b>Figura 7.13</b> Análise da evolução semanal por código postal obtida com o Saiku.....	63
<b>Figura 7.14</b> Evolução semanal em Gráfico de Barras por Código Postal com o Saiku	64
<b>Figura 7.15</b> Dashboard desenvolvido com o CDE .....	65

## ÍNDICE DE TABELAS

<b>Tabela 6.1</b> Comparação Plataformas Open Source BI.....	45
<b>Tabela 6.2</b> Pesos atribuídos a cada característica na categoria Funcionalidade .....	47
<b>Tabela 6.3</b> Pesos atribuídos a cada categoria.....	48
<b>Tabela 6.4</b> Pontuação na categoria funcionalidade.....	49
<b>Tabela 6.5</b> Pontuação final da OSSpal.....	50

## **Lista de abreviaturas, acrónimos e siglas**

<b>B2C</b>	– <i>Business to Consumer</i>
<b>BI</b>	– <i>Business Intelligence</i>
<b>CDE</b>	– <i>Community Dashboard Editor</i>
<b>DSA</b>	– <i>Data Staging Area</i>
<b>DW</b>	– <i>Data Warehouse</i>
<b>ERP</b>	– <i>Enterprise Resource Planning</i>
<b>ETL</b>	– <i>Extract, Transform and Load</i>
<b>FCS</b>	– Fator Crítico de Sucesso
<b>FLOSS</b>	– <i>Free Livre Open Source Software</i>
<b>GEO/GIS</b>	– <i>Geographic Information System</i>
<b>GM</b>	– <i>Good Mailman</i>
<b>HOLAP</b>	– <i>Hybrid OLAP</i>
<b>HTML</b>	– <i>Hypertext Markup Language</i>
<b>KPI</b>	– <i>Key Performance Indicator</i>
<b>MDX</b>	– <i>Multidimensional Expressions</i>
<b>MOLAP</b>	– <i>Multidimensional OLAP</i>
<b>OLAP</b>	– <i>On-Line Analytical Processing</i>
<b>OLTP</b>	– <i>On-Line Transactional Processing</i>
<b>OMM</b>	– <i>Open Source Maturity Model</i>
<b>Open BQR</b>	– <i>Open Business Quality Rating</i>
<b>OpenBRR</b>	– <i>Open Business Readiness Rating</i>
<b>OSMM</b>	– <i>Open Source Maturity Model</i>
<b>PDF</b>	– <i>Portable Document Format</i>
<b>PME</b>	– Pequena e Média Empresa
<b>PSW</b>	– <i>Pentaho Schema Workbench</i>

<b>QSOS</b>	– <i>Qualification and Selection of Open Source Software</i>
<b>ROLAP</b>	– <i>Relational OLAP</i>
<b>SGBD</b>	– Sistema de Gestão de Bases de Dados
<b>SQL</b>	– <i>Structured Query Language</i>
<b>SQO-OSS</b>	– <i>Software Quality Observatory for Open Source Software</i>
<b>SVG</b>	– <i>Scalable Vector Graphics</i>
<b>TMV</b>	– Tempo Médio de Vida
<b>XML</b>	– <i>eXtensible Markup Language</i>

## 1 INTRODUÇÃO

O contexto empresarial atual é cada vez mais competitivo. A capacidade das organizações responderem de forma pronta às alterações de mercado e se auto ajustarem por forma a se manterem competitivas é um fator crítico de sucesso. Isto significa que as organizações necessitam de agilidade na tomada de decisões, quer a nível estratégico, quer a nível tático ou operacional (Vercellis, 2009).

Quando o acesso a informação estruturada e o seu tratamento informático era reduzido, as tomadas de decisão envolviam um grau considerável de intuição. Atualmente esse grau pode ser reduzido ao mínimo e a tomada de decisão pretende-se que seja fortemente fundamentada em informação credível.

Deste modo, a informação produzida e obtida dentro das organizações assume-se, cada vez mais, como um ativo de elevada importância e valor, sendo até apelidada de “novo petróleo”. Uma organização que dispõe de dados de qualidade, credíveis e oportunos, encontra-se em vantagem competitiva relativamente a uma parte dos seus pares. O tratamento destes dados e transformação em informação e conhecimento, dar-lhe-á vantagem competitiva adicional.

Esta realidade leva as organizações a sentirem a necessidade de dispor de meios que facilitem a aquisição, o processamento e a análise de grandes quantidades de dados, e que sirvam como uma base sólida para descoberta de novo conhecimento (Olszak & Ziemba, 2007). Assim, o desenvolvimento de sistemas que permitam realizar análises para tomada de decisão são, cada vez mais, identificados como essenciais na melhoria da quantidade e qualidade da informação disponível para a tomada de decisão nas organizações (Turban, Sharda, & Delen, 2011). É neste contexto que surgem conceitos como o *Business Intelligence* (BI) e sistemas que o suportam.

Os sistemas de *Business Intelligence* assumem atualmente uma importância eminente dentro das organizações, disponibilizando informação útil ao processo de tomada de decisão e potenciando a qualidade da mesma. Segundo Negash (2004), *Business Intelligence* pode ser definido como um conjunto de aplicações e tecnologia para recolha, armazenamento e análise de dados com o objetivo de proporcionar melhores tomadas de decisão.

Um dos objetivos do *Business Intelligence* é a otimização de recursos e processos e, conseqüentemente, custos. Segundo Madewell (2014), o BI pode de facto trazer um



grande retorno de investimento e melhorias operacionais, mas pode também ter custos potencialmente altos. Se no início existia um domínio por parte das aplicações comerciais, que, por si só, representavam um custo normalmente elevado apenas com licenciamento, tal já não é completamente verdade hoje em dia. As soluções *open source* têm ganho expressão de uma forma geral e na área do *Business Intelligence* não é diferente, sendo atualmente uma alternativa viável às soluções comerciais, sem custos de licenciamento.

Neste contexto, a motivação principal para a realização deste trabalho foi evidenciar a relevância do tema de *Business Intelligence* na atualidade, bem como comprovar que as soluções *open source* de *Business Intelligence* como uma alternativa viável às soluções comerciais.

## **1.1 Objetivos deste trabalho**

Com este trabalho pretendemos investigar e analisar as soluções *open source* de BI disponíveis no mercado, contribuindo para a divulgação das mesmas. Posteriormente, pretendemos compará-las e avaliá-las com o objetivo de identificar a que atualmente se apresenta como mais vantajosa para implementação numa organização. Por fim, efetuamos uma demonstração prática, por meio de um estudo de caso, da viabilidade e relativa simplicidade de implementação e uso de uma solução *open source* de BI. Com este estudo de caso, pretendemos implementar localmente, como prova de conceito, a solução identificada anteriormente e com recurso a informação e dados de uma empresa real.

De um modo geral, os objetivos deste trabalho são:

- Apresentar a relevância atual do *Business Intelligence*;
- Analisar as soluções *open source* de BI como uma alternativa viável às soluções comerciais;
- Comparar e avaliar as principais plataformas *open source* de BI disponíveis atualmente;
- Identificar a plataforma *open source* de BI que apresenta maior potencial para implementação na atualidade;
- Demonstrar na prática a utilização da plataforma identificada, recorrendo a informação e dados reais de uma empresa.

## **1.2 Metodologia**

A realização deste trabalho assenta em duas fases: uma de investigação e outra de desenvolvimento.

A fase de investigação é composta inicialmente por uma revisão de literatura, com o objetivo de enquadrar e clarificar os vários conceitos que se associam ao *Business Intelligence*, seguida por uma análise do estado da arte relativamente ao estudo de plataformas *open source* de BI.

De seguida, são descritas e comparadas as três principais plataformas *open source* de BI. Com recurso a uma metodologia de avaliação de *software open source*, são avaliadas as duas plataformas com melhor resultado na comparação anterior. O objetivo é selecionar uma para utilização na fase de desenvolvimento.

A fase de desenvolvimento é constituída por um estudo de caso onde são aplicados processos e conceitos de *Business Intelligence*. É implementada localmente a plataforma identificada anteriormente com recurso a informação e dados de uma empresa real.

## **1.3 Contribuições deste trabalho**

Relativamente a contribuições, no decorrer deste trabalho, no processo de escolha da plataforma com mais potencial, foram elaborados dois artigos que foram submetidos, aceites e apresentados em conferências:

- “Avaliação Comparativa de Plataformas de *Business Intelligence Open Source* para PME”, CISTI – 13<sup>a</sup> Conferência Ibérica de Sistemas e Tecnologias de Informação, Cáceres, Espanha, Junho 2018 (Anexo 1)
- “*Open Source Business Intelligence Platforms’ Assessment using OSSpal Methodology*”, ICE-B – *International Conference on e-Business*, Porto, Portugal, Julho 2018 (Anexo 2).

Para além da elaboração dos artigos anteriormente mencionados, este trabalho contribui para o aumento da consciência coletiva relativamente à crescente importância do *Business Intelligence* em todas as áreas da sociedade. Contribui ainda para um esclarecimento relativamente às soluções *open source* em *Business Intelligence* disponíveis atualmente, bem como para as suas vantagens em termos de custos e acessibilidade em termos de implementação.

## **1.4 Estrutura do trabalho**

Este trabalho encontra-se dividido em oito capítulos, sendo o capítulo atual, Introdução, aquele onde são apresentadas as motivações e os objetivos que se pretendem atingir com a realização deste trabalho. Além disso são descritas as principais contribuições e a estrutura do trabalho.

No capítulo 2 são apresentados os conceitos de *Business Intelligence* e sistemas de *Business Intelligence*, centrais no desenvolvimento deste projeto. Explica-se a importância do *Business Intelligence* na tomada de decisão dentro das organizações e como este pode ajudar os agentes decisores neste processo. Em seguida apresenta-se uma arquitetura geral de um sistema de *Business Intelligence*, sendo posteriormente explicados os conceitos associados ao mesmo, de forma a facilitar o entendimento do sistema no seu todo e em cada uma das partes que o constituem.

No capítulo 3 são apresentados alguns trabalhos de outros autores cuja temática se relaciona com a desenvolvida neste projeto. Ter presente o que já foi feito na área ajuda-nos a orientar o nosso trabalho para algo novo e relevante a nível científico, ao mesmo tempo que permite perceber como algumas das plataformas foram evoluindo (ou não) ao longo do tempo.

No capítulo 4 são apresentadas as metodologias de avaliação de *software open source* e explorada mais em pormenor a utilizada neste projeto para avaliar plataformas *open source* de BI.

No capítulo 5 são exploradas as principais plataformas *open source* de BI existentes no mercado atualmente, sendo descritas as principais características de cada uma.

No capítulo 6 é abordado todo o processo desenvolvido para a seleção da plataforma a utilizar na demonstração prática. Na primeira parte são comparadas as plataformas em estudo tendo em conta os critérios da Gartner atuais. Numa segunda parte, as duas plataformas mais relevantes são submetidas a uma avaliação segundo a OSSpal, uma metodologia adequada e previamente selecionada e descrita, de forma a ser escolhida a plataforma recomendada atualmente para implementação e a utilizar na demonstração prática.

O capítulo 7 leva a efeito uma demonstração prática da implementação de uma plataforma *open source* de BI com recurso a dados reais de uma empresa portuguesa. É elaborada a caracterização da empresa e do negócio bem como é efetuado um levantamento de

requisitos com objetivos a cumprir no final da implementação. São aplicados processos e conceitos de *Business Intelligence* durante esta demonstração prática.

Por fim, no capítulo 8 são apresentados os resultados e conclusões do trabalho realizado, sendo também evidenciadas as dificuldades sentidas no desenvolvimento do mesmo. São apresentadas também algumas propostas de trabalho futuro.



## **2 BUSINESS INTELLIGENCE**

O termo *Business Intelligence* (BI) terá sido introduzido por Howard Dresner do Grupo Gartner em 1989 (Power, 2007). Este, segundo Power (2007), define-o como um conjunto de conceitos e métodos que pretendem melhorar a atividade de tomada de decisão de negócio, apoiada pelos sistemas de suporte à decisão baseados em factos.

O termo *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 dados em conhecimento (Negash, 2004).

Para Sezões, Oliveira, & Baptista (2006), *Business Intelligence* é um conceito que engloba um vasto conjunto de aplicações de apoio à tomada de decisão que possibilitam um acesso rápido, partilhado e interativo da informação, bem com a sua análise e manipulação, permitindo transformar grandes quantidades de informação em conhecimento útil e descobrir relações e tendências.

*Business Intelligence* pode ser visto como um termo agregador que engloba arquiteturas, bases de dados, ferramentas analíticas, aplicações e metodologias (Turban et al., 2011).

Por sua vez, Habul & Pilav-Velic (2010), definem *Business Intelligence* como um processo de recolha, tratamento, análise e utilização de informação estratégica para as organizações.

Nas últimas décadas, tem-se verificado um crescimento acentuado do número de serviços de *Business Intelligence*, bem como da adoção destes por parte das organizações (Chaudhuri, Dayal, & Narasayya, 2011), o que tem feito com que a área de *Business Intelligence* tenha vindo cada vez mais a ser identificada como essencial para a melhoria da quantidade e qualidade da informação disponível para a tomada de decisão nas organizações (Turban et al., 2011). Os sistemas de Business Intelligence representam para os gestores um apoio importante na gestão da informação dentro da sua organização. A utilização destes sistemas permite aos gestores tomar decisões menos intuitivas e mais fundamentadas.

Um sistema de BI é definido por Davenport (2006) como um conjunto de processos e de *software* utilizado para recolher, analisar e disseminar dados, com o intuito de uma melhor tomada de decisão.

Segundo Santos & Ramos (2009), os sistemas de *Business Intelligence* ajustam a recolha de dados operacionais, permitem o seu posterior armazenamento em repositórios adequados que, por sua vez, vão permitir a gestão de conhecimento através de diferentes ferramentas de análise, exploração e apresentação da informação dita essencial para a tomada de decisão.

De um modo geral, um sistema de *Business Intelligence* serve para (Olszak & Ziemba, 2007; Santos & Ramos, 2009):

- Utilizar dados históricos, desempenhos passados e atuais para elaborar previsões;
- Prever fenómenos e tendências;
- Perceber mudanças por análise e comparação de dados passados com dados atuais;
- Responder a questões não predefinidas através de acessos *ad-hoc* à informação;
- Obter um conhecimento profundo das atividades da organização.

O *Business Intelligence* permite visualizar a organização como um todo, explorando os processos de negócio de forma a que os mesmos possam ser usados como vantagem competitiva. Desta forma, os gestores têm acesso a informação sobre o mercado, a concorrência, os clientes, os processos de negócio e a tecnologia a fim de antecipar mudanças e ações da concorrência (Sassi, 2010).

Segundo Gaardboe & Svarre (2017), existem alguns fatores que podem ser determinantes no sucesso da implementação e aplicação de sistemas de *Business Intelligence*. Estes fatores críticos de sucesso (FCS) podem ser divididos em quatro categorias: tarefas, pessoas, estrutura e tecnologia. Na categoria “tarefas” referem que o FCS é o ajuste entre tarefas e tecnologia: se esta conjugação existir potencia uma elevada eficiência. Na categoria “pessoas” o FCS mais relevante é a experiência que o utilizador tem com tecnologias: quanto maior for a sua experiência, maior será a sua perceção de utilidade e facilidade de uso. Na categoria “estrutura” apontam como FCS o suporte por parte da administração: este reflete o nível a que a administração apoia os sistemas de informação, seja como patrocinador ou promotor, seja pela disponibilidade em alocar recursos aos sistemas de informação. Na categoria “tecnologia” o FCS mais relevante é a qualidade do

sistema: esta deverá ser assegurada através de uma gestão efetiva de dados e acesso à sua fonte. Além disso o modelo de dados deverá ser flexível e escalável.

Uma implementação com sucesso resulta em vários benefícios. Segundo Madewell (2014), a implementação de um sistema de BI pode trazer um grande retorno de investimento e melhorar a eficiência operacional. Operar de forma mais eficiente traduz-se num melhor serviço ao cliente que por sua vez constitui uma vantagem competitiva (Madewell, 2014).

Contudo, devemos ter sempre presente a possibilidade de nos depararmos com barreiras e dificuldades durante o processo de implementação do sistema. Segundo Madewell (2014), indica os custos com *hardware*, *software*, implementação e pessoal como potencialmente altos e com o risco de não compensarem como uma das desvantagens. Assim, as implementações de BI devem ser cuidadosamente consideradas e planeadas antes de começarem (Madewell, 2014).

Apesar de tudo, devemos ter sempre presente a importância de um sistema destes dentro de uma organização e que a sua implementação “*pode contribuir de forma eficiente e precisa no desenvolvimento da mesma, criando o conhecimento necessário para projetos futuros, de modo a minimizar falhas do passado e maximizar o seu desempenho*” (Costa, 2012).

## **2.1 Importância do *Business Intelligence* na tomada de decisão**

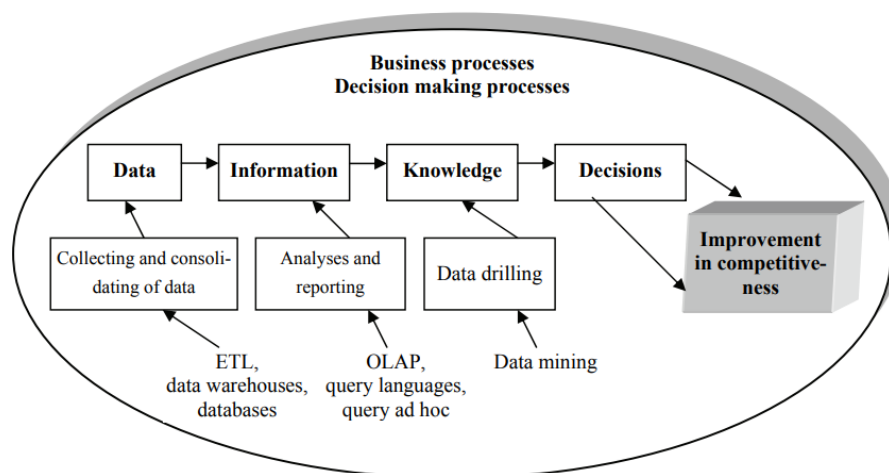
Segundo Vercellis (2009), o processo de tomada de decisão é o processo pelo qual o sujeito tenta preencher uma falha ou aproveitar uma oportunidade nas condições operacionais atuais de um determinado sistema (como ele é), e as condições mais vantajosas a serem alcançadas no futuro (o que vai ser).

A forma como decorre o processo de tomada de decisão depende das características de uma organização, das atitudes dos gestores que tomam as decisões, da disponibilidade de metodologias adequadas para a resolução de problemas, e da disponibilidade de tecnologias eficazes que suportem e apoiem a tomada de decisão (Vercellis, 2009).

Para tomar decisões corretas e no momento oportuno, os gestores necessitam de ter acesso a sistemas que contenham informação com qualidade, relativa às diversas áreas organizacionais; isto possibilita que o seu processo de decisão seja mais eficiente.



Os sistemas de BI são responsáveis pela transformação de dados em informação e desta em conhecimento, criando ambiente para uma tomada de decisão eficaz, pensamento estratégico e atuação nas organizações (ver Figura 2.1) (Olszak & Ziemia, 2007). Estes sistemas trabalham a informação, os conhecimentos, a comunicação aberta e a partilha de conhecimentos, juntamente com uma abordagem holística e analítica dos processos de negócio (Olszak & Ziemia, 2007).



**Figura 2.1** Papel do BI na tomada de decisão

**Fonte:** Olszak & Ziemia (2007, p.137)

Os sistemas de BI podem ser utilizados para suporte à tomada de decisão em todos os níveis de gestão, independentemente do seu nível de estruturação (Olszak & Ziemia, 2007):

- A nível estratégico, tornam possível estabelecer objetivos com precisão e acompanhar a realização de tais objetivos; permitem realizar diferentes relatórios comparativos, como por exemplo, os resultados históricos, a rentabilidade de uma determinada oferta, a eficácia dos canais de distribuição, juntamente com a realização de simulações de desenvolvimento ou previsão de resultados futuros com base em algumas suposições;
- A nível tático, podem fornecer alguma base para a tomada de decisões dentro do marketing, vendas, finanças, gestão de capital, etc. Permitem ainda otimizar ações futuras, modificando aspetos organizacionais, financeiros ou tecnológicos do desempenho da empresa, de modo a que esta alcance os seus objetivos estratégicos de maneira mais eficaz;

- A nível operacional, são usados para realizar análises *ad-hoc* e responder a questões relacionadas com as operações em tempo real nos departamentos, a questões da situação financeira atual, a questões de vendas e cooperação com fornecedores e/ou clientes, etc.

A importância do *Business Intelligence* para as organizações está sobretudo relacionada com o facto destes sistemas terem a capacidade de realçarem informações que podem servir de base à realização de mudanças fundamentais numa organização, as quais podem resultar no estabelecimento de novas cooperações, na aquisição de novos clientes, na criação de novos mercados, na oferta de novos produtos aos clientes (Chaudhary, 2004; Olszak & Ziemba, 2004; Reinschmidt & Francoise, 2000) em (Olszak & Ziemba, 2007).

## **2.2 Arquitetura de sistemas de *Business Intelligence***

É possível identificar diferentes arquiteturas de sistemas de *Business Intelligence* diferentes entre si, de acordo com cada autor e contexto em que estavam a ser aplicadas (Al-Debei, 2011; Chaudhuri & Dayal, 1997; Chaudhuri et al., 2011; Han, Kamber, & Pei, 2012).

Tendo por base a arquitetura apresentada por Chaudhuri et al. (2011), a Figura 2.2 apresenta uma arquitetura geral de um sistema de *Business Intelligence*, mostrando, de forma simples, a integração de cada elemento que o compõe.

Na Figura 2.2 é possível visualizar a infraestrutura dividida em cinco ambientes diferentes (Costa, 2012): Fonte de dados, Movimentação de Dados, *Data Warehouse*, *Mid-Tier* e *Front-end*.

No ambiente de fontes de dados encontram-se todas as origens dos dados que vão suportar o sistema. As fontes de dados poderão ser internas e/ou externas à organização, como *Enterprise Resource Planning* (ERP), folhas de cálculo, ficheiros de texto, bases de dados isoladas, entre outras.

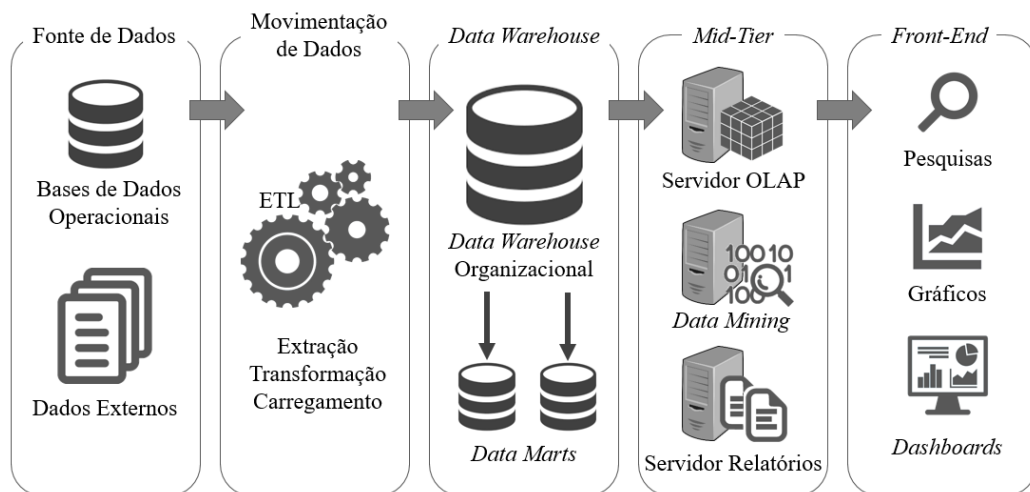
No compartimento seguinte encontra-se o ambiente de movimentação de dados. Aqui realiza-se todo o processo de *Extract, Transform and Load* (ETL), utilizando ferramentas apropriadas para o tratamento dos dados, que provêm das diversas fontes, de forma a integrar os mesmos, transformando-os e carregando-os para os repositórios de dados apropriados.

O ambiente seguinte é o ambiente de *Data Warehouse*. Este integra o *Data Warehouse* e os diversos *Data Marts* da organização, que são repositórios de informação organizacional. Estes são normalmente carregados e atualizados no processo de ETL a partir das várias fontes de dados existentes na organização.

No penúltimo compartimento encontra-se o ambiente de servidores *mid-tier*. Aqui é possível trabalhar os dados, acedendo ao *Data Warehouse* ou *Data Marts* organizacionais, com recurso a várias técnicas, como *On-line Analytical Processing* (OLAP) e *Data Mining*, de forma a poder-se gerar informação relevante para a tomada de decisão e disponibilizar a mesma aos gestores.

No último compartimento temos o ambiente de análise de negócio, onde existem várias aplicações de *front-end* que permitem o acesso e a manipulação da informação, para que os gestores possam acompanhar o desempenho do negócio utilizando ferramentas como *dashboards* e relatórios com os resultados obtidos.

Depois desta análise percebe-se que existe um conjunto grande de tecnologias associadas a um sistema de *Business Intelligence*. É importante por isso explorar alguns desses conceitos para melhor entendermos o que cada um deles significa e representa.



**Figura 2.2** Arquitetura geral de um sistema de *Business Intelligence*  
**Fonte:** Adaptado de Chaudhuri et al. (2011)

### 2.3 Conceitos associados a um Sistema de *Business Intelligence*

Para permitir uma melhor compreensão do sistema de *Business Intelligence* iremos nesta secção enumerar e explanar alguns conceitos importantes. Serão abordados os conceitos de Dados Operacionais, ETL, *Data Warehouse*, Modelo Multidimensional, OLAP (*On-line Analytical Processing*), *Data Mining* e *Dashboard*.

### **2.3.1 Sistema de Dados Operacionais**

O pré-requisito fundamental para se poder implementar sistemas de *Business Intelligence* numa empresa é esta possuir dados operacionais armazenados, uma vez que estes representam a matéria prima para um projeto deste género. Estes dados têm origem no normal processo operacional da empresa e são registados por sistemas operacionais OLTP (*On-Line Transactional Processing*) que existem na organização, concebidos para registar as operações que ocorrem diariamente na mesma. No entanto estes sistemas não estão otimizados para a análise integrada dos dados. Segundo Chaudhuri & Dayal (1997), estes registos diários são estruturados, repetitivos e consistem em transações atómicas e isoladas.

### **2.3.2 Processos ETL**

Para que seja possível criar um repositório de dados (*Data Warehouse*) preparado especialmente para ser utilizado na tomada de decisão na organização, os dados devem ser extraídos dos sistemas OLTP, devem ser transformados e só depois carregados para a *Data Warehouse* (Ariyachandra & Watson, 2010). Este processo é o processo ETL e inclui um conjunto de ferramentas especializadas de extração, transformação e carregamento, que permitem realizar o tratamento dos dados e a sua homogeneização, tendo em conta a sua complexidade, a sua limpeza e o respetivo carregamento (Vassiliadis, Simitsis, & Skiadopoulos, 2002).

As extrações dos dados podem ser distinguidas numa primeira extração, que corresponde ao primeiro carregamento dos dados para o *Data Warehouse* vazio, e as posteriores extrações incrementais dos dados que vão atualizar o *Data Warehouse* com os novos dados que vão surgindo gradualmente, da atividade diária da organização ao longo do tempo. A esta segunda extração dá-se o nome de atualização do *Data Warehouse* (Santos & Ramos, 2009).

Estando os dados devidamente extraídos, algumas transformações e limpezas são necessárias de forma a melhorar a sua qualidade (Han et al., 2012). Assim, os dados são normalmente propagados temporariamente para a área de estágio dos dados (*Data Staging Area* - DSA) onde são transformados e limpos, antes de serem carregados para o *Data Warehouse* (Santos & Ramos, 2009). É nesta fase que as anomalias habitualmente encontradas nos dados (duplicação de dados, diferentes representações de valor, dados com erros, dados em falta) são removidas (Han et al., 2012; Vercellis, 2009). A atividade

de transformação e limpeza permite que os dados se apresentem num formato homogéneo, consistente e isentos de erros. (Santos & Ramos, 2009).

Depois de concluído todo o processo de transformação, temos o processo de carregamento em que os dados que se encontram na DSA são carregados para o *Data Warehouse* (Santos & Ramos, 2009).

### **2.3.3 Data Warehouse**

O termo *Data Warehouse*, usado pela primeira vez em 1991 por Inmon, é definido pelo mesmo como sendo um conjunto de dados integrados, orientados e organizados a um determinado assunto, não voláteis e estruturados temporalmente de forma a suportar o processo de tomada de decisão (Inmon, 2005).

Segundo a perspetiva de Inmon, uma *Data Warehouse* é caracterizada por ser (Inmon, 2005):

- **Organizada e orientada por assunto/tema:** os dados são organizados por assuntos/temas chave de uma organização. São normalmente apresentados de forma compartimentada, de acordo com as necessidades específicas dos utilizadores finais. Desta forma a sua análise será facilitada e resultará numa visão simples de um determinado assunto;
- **Integrada:** a *Data Warehouse* apresenta-se como uma fonte de dados única e ao mesmo tempo abrangente sobre e para o negócio. Os vários dados que se encontram nas diversas fontes de dados são selecionados, integrados e posteriormente armazenados numa *Data Warehouse*, garantindo e fornecendo uma visão única e coerente do negócio, de forma a responder às questões que são colocadas pelos utilizadores;
- **Catalogada temporalmente:** a *Data Warehouse* apresenta os dados do passado e a informação atual sobre o negócio. O seu objetivo principal é fornecer informação válida sobre a perspetiva histórica;
- **Não volátil:** na *Data Warehouse* podem ser adicionados novos dados, mas não devem ser removidos os dados lá armazenados. Por isso, no ambiente de *Data Warehouse* existem três tipos de operações: o carregamento inicial dos dados, a posterior atualização dos dados e o acesso aos dados para processamento de consultas.

Han et al. (2012), consideram a *Data Warehouse* como sendo um repositório de dados consistentes, através do qual se compõe um modelo de dados de suporte à decisão, armazenando informação relevante para a tomada de decisão estratégica da organização.

Por isso, uma *Data Warehouse* deve (Kimball & Ross, 2002):

- Permitir fácil acesso à informação da organização: o conteúdo deve ser compreensível, de modo, a informação ser intuitiva e óbvia para o utilizador.
- Apresentar a informação de forma consistente: a informação deve ser credível. Se existem duas medidas de desempenho iguais então devem significar a mesma coisa, por outro lado, se duas medidas não têm o mesmo significado, têm de ser medidas de desempenho diferentes.
- Ser adaptável e resistente a mudanças: as mudanças não podem ser simplesmente evitadas, a *Data Warehouse* deve ser projetada para lidar com as mudanças.
- Proteger a informação da organização: normalmente uma *Data Warehouse* armazena informação de extrema importância para uma organização. Desse modo, uma *Data Warehouse* deve ter um controlo efetivo sobre o acesso à informação.
- Servir de base para a melhor tomada de decisão: a *Data Warehouse* deve ter os dados corretos para apoiar a tomada de decisão.
- Ser aceite pela comunidade organizacional: as diferentes áreas de negócio devem adotar a *Data Warehouse*.

Uma *Data Warehouse* deve integrar informação referente a um determinado assunto, ou vários, da organização, caracterizando-a como um todo e não parte dela. Se uma *Data Warehouse* integrar informação de âmbito mais restrito, por exemplo, apenas de um departamento, armazenando um subconjunto de dados específico da organização, designa-se por *Data Mart* e não *Data Warehouse* (Santos & Ramos, 2009).

As *Data Marts* são repositórios de dados multidimensionais, mais pequenos do que as *Data Warehouses*, que reúnem um conjunto de tabelas dimensionais de suporte a um determinado processo de negócio (Kimball & Ross, 2002).

### 2.3.4 Modelo Multidimensional

A organização dos dados numa *Data Warehouse* ou *Data Mart* é feita de acordo com o modelo multidimensional.

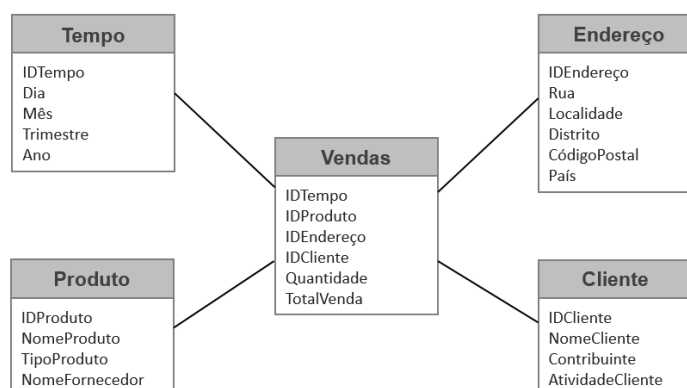
Neste modelo, os dados são organizados como um cubo, sendo este definido segundo factos, dimensões e hierarquias (Golfarelli, Maio, & Rizzi, 1998). Esta modelação pode ser implementada segundo um dos seguintes esquemas (Chaudhuri & Dayal, 1997):

- Esquema em Estrela (*Star Schema*);
- Esquema em Constelação (*Constellation Schema*);
- Esquema em Floco de Neve (*Snowflake Schema*).

O esquema em estrela é a forma mais comum de modelar uma *Data Warehouse*. Este esquema (ver Figura 2.3) integra uma tabela de factos e várias tabelas de dimensões.

As tabelas de factos correspondem ao assunto que se pretende analisar, normalmente uma componente de negócio (por exemplo, vendas, compras, movimentos de stock). Estas tabelas ocupam a maior parte do espaço ocupado pela *Data Warehouse* (Santos & Ramos, 2009).

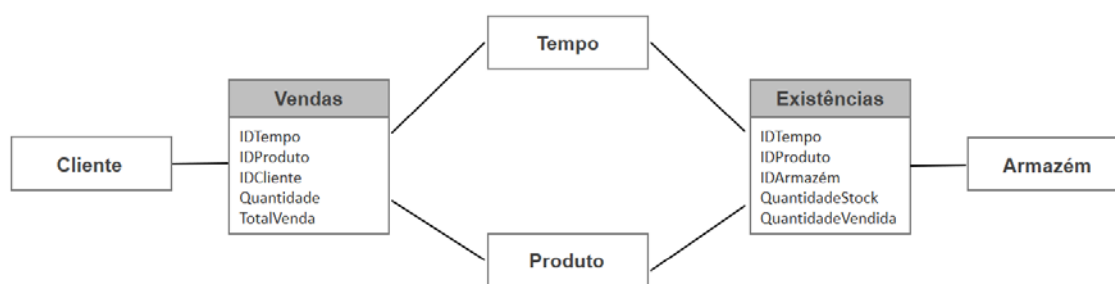
As tabelas de dimensão correspondem às variáveis de análise, pelas quais a tabela de factos pode ser analisada (por exemplo, tempo, produto, clientes). Estas tabelas são geralmente normalizadas e estruturadas hierarquicamente (por exemplo, [ano, mês, dia]) (Vercellis, 2009).



**Figura 2.3** Exemplo de Esquema em Estrela

**Fonte:** Santos & Ramos (2009)

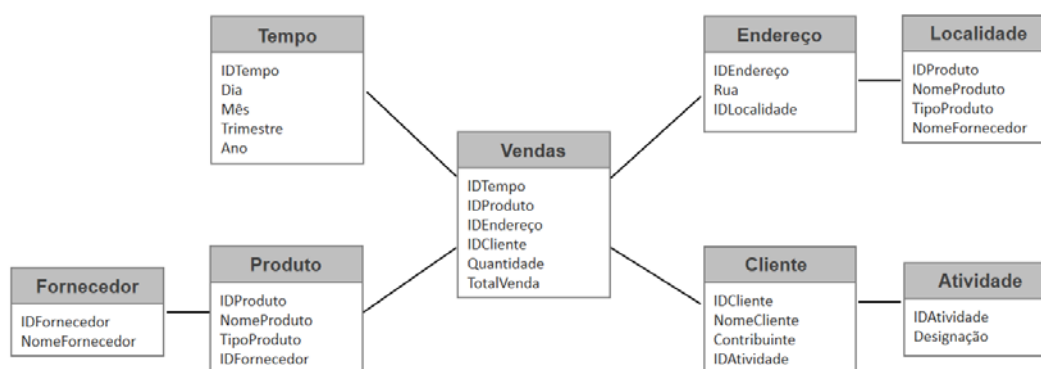
O esquema em constelação (ver Figura 2.4) é um modelo de dados mais complexo, onde várias tabelas de factos partilham uma ou mais tabelas de dimensão (Santos & Ramos, 2009).



**Figura 2.4** Exemplo de Esquema em Constelação

**Fonte:** Santos & Ramos (2009)

O esquema em floco de neve (ver Figura 2.5) é semelhante ao esquema em estrela, mas onde todas as suas tabelas de dimensão se encontram normalizadas. Apesar destes dois esquemas serem equivalentes em termos de conteúdo de dados, o esquema em floco de neve acaba por apresentar uma estrutura mais complexa, o que torna, por vezes, a sua compreensão mais complicada (Santos & Ramos, 2009).



**Figura 2.5** Exemplo de Esquema em Floco de Neve

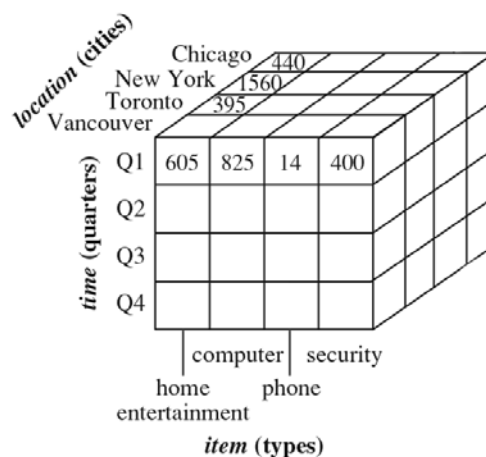
**Fonte:** Santos & Ramos (2009)

### 2.3.5 Sistemas OLAP (*On-line Analytical Processing*)

Kimball & Ross (2013) definiram OLAP como um conjunto de princípios que fornecem uma *framework* multidimensional para o apoio à decisão, com modelos dimensionais implementados em bases de dados multidimensionais.

Utilizando estes sistemas é possível criar e gerir cubos multidimensionais (ver Figura 2.6), que possibilitam a análise de informação sob diferentes perspetivas, cuja construção está facilitada pelo próprio modelo da *Data Warehouse* ser multidimensional. As análises que recorrem a este tipo de sistemas são baseadas em hierarquias de conceitos para consolidar os dados e para criar visualizações ao longo das dimensões de uma *Data Warehouse* (Vercellis, 2009).





**Figura 2.6** Exemplo de um cubo de 3 dimensões  
**Fonte:** Kamber et al. (2011)

Os servidores OLAP, tal como já foi referido, permitem a análise multidimensional dos dados. Estes sistemas permitem ao utilizador realizar, num ambiente amigável, operações de manuseamento e análise interativa sobre os cubos (Santos & Ramos, 2009) (ver Figura 2.7):

- **Drill-down** ou **Roll-down**: esta operação permite navegar desde os dados mais generalizados até aos dados mais detalhados. Desta forma consegue-se uma visão mais pormenorizada dos dados a analisar;
- **Roll-up** ou **Drill-up**: esta operação permite analisar inversamente os dados, isto é, desde os dados mais pormenorizados até aos mais agregados. Operação oposta do *Drill-down*;
- **Slice and Dice**: esta operação permite limitar a visualização de informação através do seu recorte (*slice*) e redução (*dice*). O recorte permite seleccionar um subconjunto de dados do cubo e estender pelas suas dimensões, a redução permite definir um subcubo através da definição de critérios de selecção para mais do que duas dimensões;
- **Pivot** ou **Rotate**: esta operação permite rodar os eixos de visualização dos dados e apresentá-los de uma forma alternativa de acordo com a rotação.

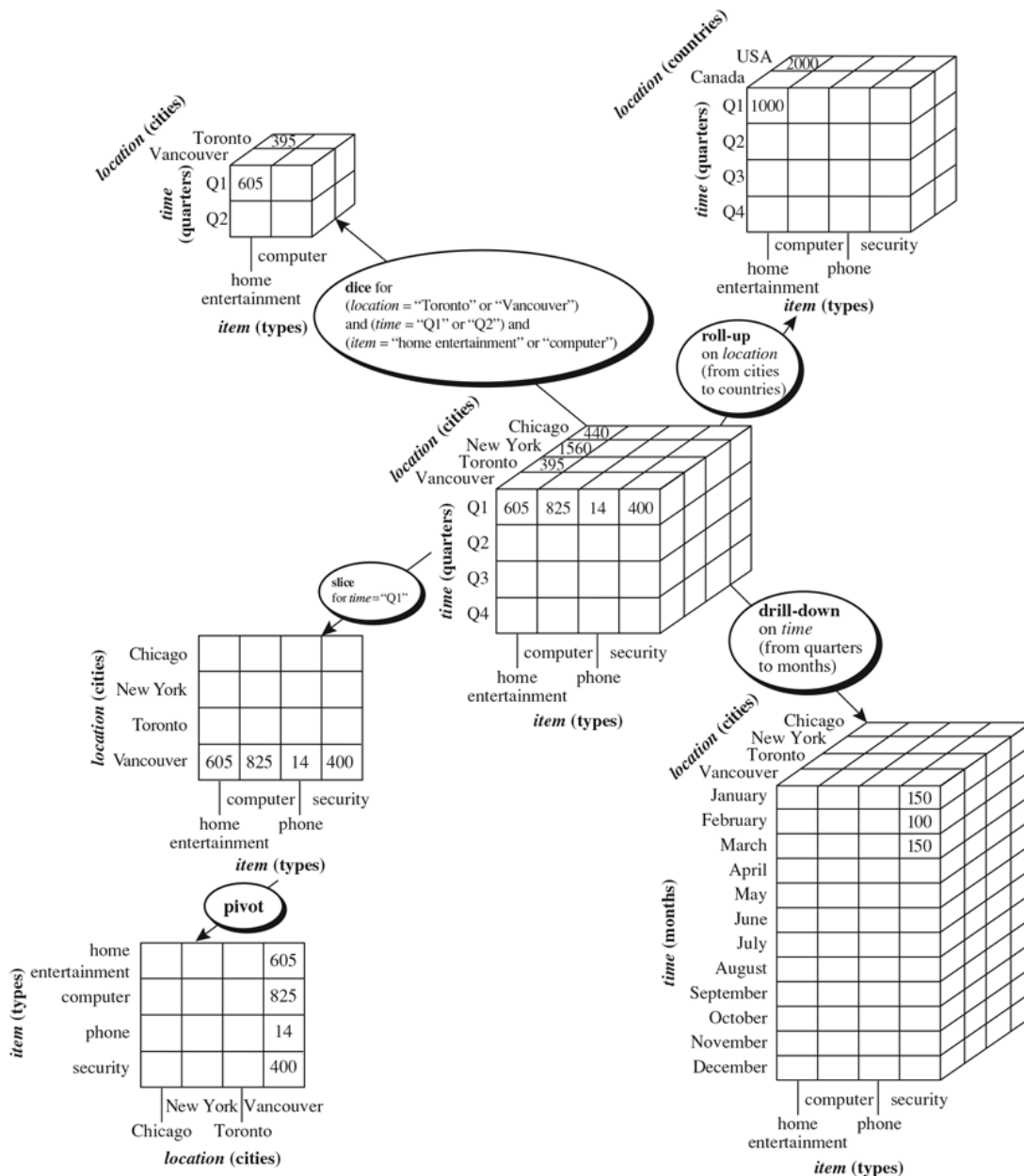


Figura 2.7 Manipulação de cubos

Fonte: Kamber et al. (2011)

Estes sistemas permitem armazenar os dados de acordo com uma das seguintes arquiteturas (Han et al., 2012; Larson, 2009; Rainardi, 2008):

- **ROLAP (Relational OLAP):** esta arquitetura atua como intermediário entre a base de dados relacional e as ferramentas de análise de dados. Utiliza um Sistema de Gestão de Base de Dados (SGBD) relacional para o armazenamento e gestão dos dados a serem analisados. Tem como vantagem a capacidade de lidar com grandes quantidades de dados; tem como desvantagem a dificuldade na recuperação dos valores agregados e

a lentidão em relação às outras arquiteturas, uma vez que cada análise efetuada representa uma ou mais consultas SQL (*Structured Query Language*) à base de dados relacional, o que no caso de esta ser grande, pode representar bastante tempo;

- **MOLAP** (*Multidimensional OLAP*): esta arquitetura, ao contrário da anterior, utiliza bases de dados multidimensionais para armazenar os dados, logo suporta vistas multidimensionais dos dados, apresentando um excelente desempenho e uma rápida indexação a dados pré-processados.
- **HOLAP** (*Hybrid OLAP*): esta arquitetura combina as duas anteriormente referidas, beneficiando dos pontos fortes de cada uma: a escalabilidade da arquitetura ROLAP e a capacidade de velocidade e maior processamento da arquitetura MOLAP.

Os sistemas OLAP enquadram-se em modelos simples de análise e identificação de padrões nos dados.

### **2.3.6 Data Mining**

Uma das partes mais importantes do BI é a exploração dos dados armazenados nas *Data Warehouses*, uma vez que os resultados desta exploração vão ajudar na tomada de decisão. É por isto que uma ferramenta como o *Data Mining*, que permite a prospeção de dados, tem um papel fundamental.

O *Data Mining* é o processo de descobrir padrões em grandes quantidades de dados. Como um processo de descoberta de conhecimento, normalmente envolve limpeza de dados, integração de dados, seleção de dados, transformação de dados, descoberta de padrões, avaliação de padrões, e apresentação do conhecimento (Han et al., 2012).

Este processo pode ter dois tipos de abordagem (Berry & Linoff, 2004): a abordagem direta ou modelo orientado à verificação; e a abordagem indireta ou modelo orientado à descoberta. A abordagem orientada à verificação é utilizada quando se sabe o que se quer pesquisar, onde o utilizador formula hipóteses, refinando-a sempre que necessário, e o sistema encarrega-se de verificar essas hipóteses (Maimon & Rokach, 2010). Contrariamente, na abordagem orientada à descoberta, os padrões nos dados são identificados de forma automática sem interferência do utilizador (Maimon & Rokach, 2010).

Para iniciar um processo de *Data Mining*, o primeiro passo é a definição de objetivos a atingir e o tipo de resultados que se pretende alcançar (Piedade, 2011). Em função do objetivo e tipo de resultado definido, é definida a tarefa a executar (classificação, segmentação, sumarização, modelação de dependências) e identificada a técnica de *Data Mining* a utilizar (árvores de decisão, regras de associação, regressão linear, redes neurais, algoritmos genéticos, entre outras). Seguidamente, aplica-se a técnica de *Data Mining* ao conjunto de dados. Pode ser necessário utilizar mais do que uma técnica de *Data Mining* para atingir os objetivos, já que a qualidade dos dados e o tipo de dados disponíveis pode influenciar os resultados obtidos (Piedade, 2011).

### **2.3.7 Dashboards**

O acompanhamento do desempenho de uma organização e o acesso aos dados pode ser feito através de diferentes aplicações. Nestas aplicações, a informação é disponibilizada de diversas formas como *dashboards*, tabelas, gráficos e ferramentas interativas de análise multidimensional.

Os *dashboards* são definidos por Few (2004) como uma representação visual da informação mais importante, necessária para atingir um ou mais objetivos, consolidada e organizada num único ecrã, de modo a que a informação possa ser monitorizada rapidamente. Permitem a apresentação de métricas de desempenho de toda a organização, tendo sempre em mente que o sucesso de uma empresa está dependente da capacidade que os seus responsáveis e colaboradores para tomar decisões acertadas em tempo útil (Malik, 2005). Para Velcu-Laitinen & Yigitbasioglu (2012), os *dashboards* são uma ferramenta interativa de desempenho visual, que permitem exibir num único ecrã a informação mais importante, de forma a que os utilizadores sejam capazes de identificar, explorar e comunicar as áreas problemáticas que necessitam ação corretiva.

Dada a sua importância, quando se constrói um *dashboard* é necessário perceber não só como se deve organizar graficamente o mesmo, mas também quais são as vistas gráficas mais apropriadas para a informação que se pretende exibir (Barros, 2013).

Para o desenvolvimento de um bom *dashboard* existem alguns requisitos (Eckerson, 2005; Few, 2006):

- **Respeitar o limite de um único ecrã** - Um dos maiores benefícios de um *dashboard* é a comunicação simultânea de informações. Informações em

ecrãs distintos podem prejudicar a análise crítica e a comparação entre indicadores;

- **Contextualizar os dados apresentados** - Algumas informações podem fornecer conclusões erradas ou perder o significado se não forem inseridas num contexto adequado;
- **Evitar a utilização excessiva de detalhes** - Fornecer informações além do necessário pode dificultar a identificação das informações mais importantes para apoiar o processo de tomada de decisão;
- **Escolher a componente gráfica apropriada** - Este requisito é um dos erros mais comuns na elaboração de um *dashboard*. A escolha da componente gráfica tem de estar relacionada com a natureza da informação e da mensagem que se pretende transmitir, nesse sentido é importante definir a componente gráfica mais adequada de modo a divulgar a informação de forma clara e eficiente, sem distrações;
- **Dispor e destacar os dados adequadamente** - Os dados devem ser organizados por ordem de importância, os quais a organização devem classificar na sequência que desejar. Os dados com maior relevância devem ser destacados dos outros;
- **Utilizar cores apropriadas** - Utilizar cores apropriadas de modo a identificar os dados que requerem mais atenção apenas quando necessário e não utilizar sempre cores fortes.

Estes são os requisitos necessários para a elaboração de um *dashboard* de forma a obter-se uma fácil visualização, interpretação dos resultados e aquisição de conhecimento.

### **3 TRABALHOS RELACIONADOS**

Neste capítulo é apresentada a revisão de literatura, sendo esta dividida em dois temas: Plataformas *Open Source* de *Business Intelligence* e Metodologias de Avaliação de *Software Open Source*.

#### **3.1 Plataformas *Open Source* de *Business Intelligence***

Thomsen & Pedersen (2008) foram dos primeiros autores a realizar um estudo científico sobre *open source* BI. No seu estudo focaram-se na comparação de diferentes tipos de ferramentas/funcionalidades de BI: três ferramentas de ETL (Bee, CloverETL e Octopus), três sistemas OLAP (Bee, Lemur e Mondrian) e quatro SGBDs (MonetDB, MySQL, MaxDB e PostgreSQL). Concluíram, nesta altura, que os SGBDs eram as ferramentas mais desenvolvidas naquele momento, aplicáveis a situações reais com grandes conjuntos de dados, apresentando um bom desempenho. Realçavam que as ferramentas de ETL *open source* ainda estavam muito pouco desenvolvidas comparadas com as ETL não *open source*, e que existiam muitas variações de desenvolvimento nos servidores OLAP. Mencionam que um dos maiores problemas para a utilização destas ferramentas *open source* é a falta de documentação adequada, que torna muitas vezes difícil perceber como uma tarefa específica pode ser executada por um determinado produto. Apesar de pioneiro, este estudo focou-se na comparação de ferramentas *open source* de BI individuais e não em plataformas de BI, que geralmente apresentam na sua arquitetura não só estas ferramentas, mas também outras não exploradas por estes autores.

Tendo em vista facilitar a escolha de uma plataforma de BI por parte das empresas, Tereso & Bernardino (2013) realizaram um estudo onde avaliaram oito grandes plataformas: quatro comerciais (IBM Cognos, Microsoft BI, MicroStrategy e Oracle BI), e quatro *open source* (JasperSoft, Pentaho, SpagoBI and Vanilla). Várias características foram consideradas na comparação: relatórios, *Dashboards*, OLAP, ETL, *Data Mining*, KPIs (*Key Performance Indicators*), *Data export*, GEO/GIS, *Ad-hoc queries*, sistemas operativos em que funcionam, versão para dispositivo móvel. Relativamente às plataformas *open source*, concluíram que a Pentaho e a SpagoBI são plataformas idênticas que possuem a maioria das características avaliadas e que podem ser utilizadas nos maiores sistemas operativos. A plataforma JasperSoft não possuía *Data Mining* ou KPIs e não suportava o sistema operativo UNIX. A plataforma Vanilla não possuía sistema GEO/GIS. Todas as plataformas comerciais eram idênticas e possuíam todas as

características usadas na avaliação comparativa. A única exceção era a Oracle que não suportava Unix. Concluíram ainda que a Pentaho e a SpagoBI eram equivalentes às plataformas comerciais, apresentando-se como uma alternativa com menores custos de aquisição.

Ferreira, Silva, Vieira, Guimarães, & Carvalho (2010) realizaram um estudo onde compararam duas plataformas de BI, uma *open source* (Pentaho) e uma comercial (Microsoft), através da sua utilização num estudo de caso. Ambas as plataformas foram utilizadas para o desenvolvimento do mesmo projeto, utilizando a mesma base de dados, e alguns critérios foram definidos e categorizados para serem utilizados na análise. Com base nos resultados obtidos, os autores concluíram que, de acordo com os diversos critérios estabelecidos, ambas as plataformas apresentavam características positivas e negativas, e que a simples nomeação de uma ser superior à outra poderia não atender a diferentes cenários e contextos. Para perceber em que situação cada uma das plataformas se apresentava como mais adequada, dividiram e organizaram os critérios avaliados de forma a perceber qual o perfil de organização para a qual a plataforma Pentaho poderia ser mais adequada, e também o perfil da organização em que a plataforma Microsoft seria a mais adequada. O primeiro perfil era referente a organizações que possuíam restrições financeiras, necessidade de adaptação funcional dos componentes e da interface da plataforma, necessidade de utilizar a plataforma em vários sistemas operativos e que permita ter acesso remoto à plataforma. Neste caso a plataforma Pentaho apresentava-se como a mais adequada. O segundo perfil era referente a organizações com recursos disponíveis para investir em BI, com baixo conhecimento na área, pouco tempo disponível para implementar e que exijam interfaces mais amigáveis e fácil utilização. Para estes casos a plataforma Microsoft mostrava-se mais adequada.

Brandão et al. (2016) realizaram uma análise comparativa de plataformas *open source* BI com a finalidade de identificar, dentro de um grupo de plataformas previamente selecionadas, qual a mais adequada ao setor da saúde. O grupo em análise era constituído pelas seguintes plataformas: QlikView, Palo BI Suite, Jaspersoft, Tableau Public, Spago BI e Pentaho BI Suite. No âmbito do setor da saúde foram identificados 23 requisitos sobre os quais foi levada a efeito a análise comparativa das ferramentas. Os 23 requisitos foram agrupados em 5 áreas e atribuída uma percentagem a cada área de acordo com a sua relevância para o sucesso da implementação. A ferramenta com melhor pontuação foi a Spago BI com 34.85 seguida da Pentaho BI Suite com pontuação quase idêntica (34.70)

e da Jaspersoft BI com 29.45 num máximo possível de 39.5. Entre a Spago BI e a Pentaho BI Suite a preterida foi a Spago BI por ter um processo de instalação consideravelmente mais complexo, pouca documentação de suporte e utilizar bastante mais RAM.

Bertolini et al. (2015) realizaram uma análise comparativa das plataformas *open source* de BI JasperSoft, Pentaho, SpagoBI e Vanilla, tendo em conta a análise da documentação disponível acerca dos mesmos. Evidenciaram que as plataformas Pentaho, SpagoBI e Vanilla possuem significativamente mais recursos, apesar das quatro plataformas em análise poderem ser executadas nos principais sistemas operacionais. Mencionam que todos apresentam vasta documentação, comunidades ativas e documentação oficial relevante para a implementação dos sistemas. Além disso, todas as comunidades de desenvolvimento apresentam serviços de consultoria pagos. Outra conclusão relevante foi o facto das plataformas SpagoBI e Vanilla apenas possuírem versão *open source*, o que foi considerado pelos autores algo positivo, uma vez que os utilizadores dessas plataformas não seriam surpreendidos por limitações de funcionalidades que estivessem restritas a versões pagas.

Ahishakiye, Omulo, Taremwa, & Wario (2017) efetuaram um estudo onde realizaram uma análise comparativa de cinco plataformas *open source* BI: Apache Hadoop, Jaspersoft, Pentaho, SpagoBI e Vanilla. As funcionalidades avaliadas foram: relatórios, *Dashboards*, OLAP, ETL, *Data Mining*, KPIs, GEO/GIS, *Ad-Hoc Queries*, compatibilidade com múltiplos sistemas operativos, Java, distribuição de armazenamento e processamento, tolerância ao erro e escalabilidade. Concluíram que a plataforma Apache Hadoop seria a plataforma recomendada pois era a que apresentava mais funcionalidades que eram consideradas importantes, seguida pela Pentaho e a SpagoBI (também com um elevado número de critérios verificados) e, depois, a Vanilla e, por fim, a Jaspersoft.

Parra & Syed (2016) realizaram uma análise experimental onde compararam duas das mais bem posicionadas plataformas *open source* BI no mercado: Pentaho e Jaspersoft. O objetivo desse trabalho foi analisar e avaliar essas plataformas e descrever como melhoram a qualidade dos dados, medindo o seu desempenho em termos de processos de ETL e *Reporting*, usando os sistemas Sage e Matlab. Segundo os autores, no que diz respeito à análise de processos de ETL, a plataforma Jaspersoft apresenta um incremento no tempo de CPU (*Central Processing Unit*) comparado com a Pentaho. Por outro lado, a plataforma Pentaho tem um incremento de tempo de CPU no processo de *Reporting*,



comparativamente com a Jaspersoft. Sendo uma análise pormenorizada a estes dois critérios, poderá ajudar na decisão de algumas empresas para as quais estes dois critérios sejam os mais importantes, e que já se encontrem em posição de decidir entre estas duas plataformas.

Golfarelli (2009) realizou uma avaliação comparativa de três plataformas: Jaspersoft, Pentaho e SpagoBI, comparando-as não só entre si, na sua versão *open source*, como também com as suas versões comerciais. Concluiu que as funcionalidades administrativas poderiam ser melhoradas nas versões *open source*, sobretudo para a plataforma Pentaho e que a plataforma SpagoBI é a que oferece mais funcionalidades sendo que a sua versão *open source* é comparável às versões comerciais da Pentaho e Jaspersoft.

Ribeiro & Bernardino (2010), realizaram uma comparação de três das plataformas *open source* BI mais utilizadas: OpenI, Pentaho e SpagoBI. Neste trabalho foram relatadas e resumidas as avaliações relativas às plataformas, não tendo sido estudadas em profundidade as características individuais de cada uma. A comparação realizada centrou-se nos seguintes aspetos-chave: interface de utilizador, recursos e usabilidade. Relativamente à interface de utilizador, a OpenI apresentava um login simples, apenas com a opção de utilizador administrador. Já a Pentaho apresenta um login mais complexo, com a possibilidade de escolha de vários utilizadores ou administradores. A SpagoBI apesar de um login tradicional, permite a escolha entre administrador ou técnico utilizador. Em termos de opções após o login, a OpenI e a Pentaho apresentam uma interface mais tradicional (*web-oriented*) com as opções apresentadas num menu à esquerda, enquanto que a SpagoBI tem uma apresentação mais tipo “ambiente de trabalho”, com o menu no topo. A OpenI difere das outras plataformas uma vez que não pode exibir a tabela de dados e o gráfico para análise em simultâneo. Todas as plataformas apresentam a opção de diversos idiomas. Em termos de recursos, os autores consideram a OpenI a mais simples das três avaliadas e a que apresenta menos recursos. As suas capacidades são focadas na análise e geração de relatórios simples. Já a Pentaho e a SpagoBI têm a capacidade de produzir relatórios mais avançados, indicadores-chave de desempenho e análise de geo-referenciamento. Em todas as plataformas, está disponível a exportação de relatórios e gráficos para outros formatos. Relativamente à usabilidade, tanto a SpagoBI como a Pentaho disponibilizam o ponto de vista da administração e o do utilizador. Já a OpenI mostra apenas a visão de administração. Nenhum teste de usabilidade formal foi realizado pelos autores, mas estes avaliaram as plataformas em

termos de execução. Na construção de documentos os autores consideraram a SpagoBI um pouco confuso, com muitos passos espalhados por vários menus. A OpenI tem um processo de criação de documentos mais simples, com uma opção de menu que permite criar novo relatório, com todas as operações a serem realizadas na mesma interface. Já a Pentaho utiliza diferentes programas para projetar e desenvolver relatórios e outros documentos. Tem um assistente que guia, passo-a-passo, o utilizador através da criação dos relatórios. Para ver um documento o utilizador tem disponível a opção “*Execute*” no menu tipo *pop-up*. Em termos globais, os autores consideraram que as plataformas *open source* BI estão em crescimento não só em termos de funcionalidades, mas também aparência e qualidade. Das três plataformas analisadas, concluíram que a OpenI é sem duvida a menos desenvolvida e que a Pentaho é a mais apelativa. Além disso, concluíram que do ponto de vista de uma possível implementação, a Pentaho deverá ser a melhor opção, uma vez que apresenta mais documentação e recursos disponíveis. Contudo, e independentemente da plataforma escolhida, é necessária destreza a trabalhar em Java e profundo entendimento da forma como funciona o modelo de *Application Server* são essenciais para implementar este tipo de *software*. Concluem também que apesar de tudo, a Pentaho não é uma “verdadeira” solução *open source* uma vez que alguns dos seus recursos mais avançados apenas se encontram disponíveis na versão comercial. Tendo isto em conta, se se procura uma solução realmente *open source* a SpagoBI deve ser a considerada.

Em 2011, Sidiqi & Mukhi realizaram uma comparação entre as plataformas Microsoft BI Suite e a Pentaho *open source*. Os autores escolheram a Microsoft BI Suite por esta ter sido considerada em diversos estudos como estando no topo do ranking. Já a escolha da Pentaho aconteceu por esta ter sido declarada pela Gartner em 2010 uma das plataformas *open source* BI emergente. Da análise das duas, os autores concluíram que ambas são boas plataformas de BI; que ambas suportam OLAP; que a plataforma da Microsoft não é compatível com Unix e Linux; que a Pentaho possui apoio para instalação, mas que este não é comparável com o melhor suporte da Microsoft; que a Pentaho oferece videos e documentação de suporte, mas esta documentação não consegue suprimir as necessidades do utilizador, enquanto que a Microsoft possui centros certificados com formadores certificados em diversos pontos; que a utilização da plataforma da Microsoft implica um custo elevado em licença enquanto que a utilização da Pentaho é gratuita. Tendo isto em conta, os autores propõem que, apesar da plataforma

da Microsoft ser mais produtiva e mais adequada para qualquer organização com boa situação financeira, a Pentaho é a melhor alternativa num momento de recessão económica.

Tereso & Bernardino (2011) realizaram um levantamento e comparação entre plataformas *open source* BI por forma a identificar a(s) mais completa(s) e adequada(s)/recomendada(s) para uso em Pequenas e Médias Empresas.

Compararam as plataformas SpagoBI, OpenI, Pentaho, JasperSoft, Palo e Vanilla, tendo em conta as suas funcionalidades, características e os sistemas operativos que as suportam. Concluíram que a ferramenta SpagoBI era a mais completa, sendo que a versão única (*community*) possuía mais funções que algumas pagas (*enterprise*). Além disso, a Spago BI era a única das plataformas analisadas com suporte para georeferenciação e versão para dispositivos móveis. Praticamente ao nível da SpagoBI estava a Pentaho que diferia por não ter suporte para georeferenciação. Esta plataforma era disponibilizada em duas versões: *community* (sem custo, menos funções) e *enterprise* (todas as funções e com custos).

As plataformas JasperSoft e Palo não apresentavam soluções para *Data Mining* e KPIs, sendo, por isso, consideradas soluções inferiores à SpagoBI, Vanilla e Pentaho. Por último concluíam que a plataforma OpenI só apresentava metade das funcionalidades da plataforma SpagoBI.

Lapa, Bernardino & Figueiredo (2014) efetuaram uma análise a 7 plataformas *open source* de BI (Actuate, Jaspersoft, OpenI, Palo, Pentaho, SpagoBI e Vanilla) segundo os critérios de Gartner 2013, aos quais consideraram relevante adicionar *Cloud* e *Collaborative Technologies*, dada a sua crescente importância junto dos grupos decisores. Verificaram que as plataformas Pentaho e Vanilla validavam 10 em 11 dos critérios avaliados e que a SpagoBI validava todos, uma vez que era a única, à data, a disponibilizar tecnologias colaborativas. Concluíram que estas 3 (Pentaho, Vanilla e SpagoBI) são competitivas relativamente a plataformas comerciais e que, embora não fosse possível identificar a melhor, tinham grande potencial de uso em Pequenas e Médias Empresas (PMEs).

Em Marinheiro & Bernardino (2015) foram comparadas 5 plataformas *open source* de BI (Jaspersoft, Palo, Pentaho, SpagoBI e Vanilla) utilizando também os critérios da Gartner 2013. Nesta comparação, destacaram-se as plataformas Pentaho e SpagoBI, as quais

foram submetidas a uma avaliação experimental com recurso à metodologia de comparação de *software open source*, OpenBRR (*Open Business Readiness Rating*). Os autores concluíram que, na escala de avaliação desta metodologia, a plataforma SpagoBI obtinha um resultado classificado como “Muito Bom”.

### **3.2 Metodologias de Avaliação de *Software Open Source***

O aumento do uso de *software Open Source* na sua versão "*Free / Libre*" *Open Source Software* (FLOSS) que temos testemunhado no final do século XX e início do século XXI, deve-se a diversos fatores incluindo a ausência de custos relacionados com a compra de licenças e à possibilidade de ter acesso a códigos fonte, o que permite aos utilizadores criarem soluções à medida do que necessitam. Uma desvantagem é a ausência de métricas que assegurem a qualidade do software e que comprovem sua validade (Petrinja, Sillitti, & Succi, 2008).

Deprez & Alexandre (2008) foram dos primeiros autores a abordar o tema. Eles conduziram uma comparação rigorosa entre a *Open Business Readiness Rating* (OpenBRR) e a *Qualification and Selection of Open Source Software* (QSOS) com base na descrição das metodologias e não na sua aplicação empírica, e identificaram vantagens e desvantagens de ambas as metodologias. Concluíram que a OpenBRR permite adaptar os critérios a um domínio, portanto, um melhor ajuste ao contexto de avaliação, mas a terminologia é ampla e imprecisa. Por outro lado, a QSOS possui uma extensa lista de critérios, mas as regras de pontuação são ambíguas para mais da metade dos critérios. Os autores concluíram ainda que a pontuação de 3 níveis (0,1 ou 2) que a QSOS disponibiliza para classificar os diferentes parâmetros de avaliação é muito restritiva.

(Petrinja, Sillitti & Succi, 2010) desenvolveram um estudo sobre a qualidade e usabilidade de três metodologias de avaliação de *software open source*: OpenBRR, QSOS e a *QualiPSo Open Source Maturity Model* (OMM). O estudo identificou os aspetos positivos e negativos de cada uma delas. Os resultados revelaram que as três metodologias forneceram avaliações semelhantes. A principal conclusão deste estudo foi que os três modelos têm parâmetros de avaliação que não têm uma formulação clara e, portanto, não são claros para os avaliadores. Em alguns parâmetros, o valor limite disponível para a pontuação também não estava claro. Os aspetos críticos de cada modelo foram: Funcionalidade e Qualidade para o OpenBRR; Adoção, administração / monitorização,

propriedade de direitos de autoria e *Browser* para QSOS, e Qualidade do Plano de Teste e o Ambiente Técnico para OMM.

Relativamente a sujeitar plataformas *open source* de *Business Intelligence* a estas metodologias de avaliação, Ferreira, Pedrosa & Bernardino (2017) avaliaram 4 plataformas *open source* de BI (Birt, Jaspersoft, Pentaho e SpagoBI) utilizando a metodologia de avaliação de *software open source* OSSpal. Aplicando a metodologia, na sua escala de avaliação (de 1 a 5), a Pentaho obteve 3.47, a SpagoBI, 2.92 e a Jaspersoft, 2.90. Por comparação com a Pentaho, a SpagoBI obteve pior desempenho na categoria de comunidade e a Jaspersoft na categoria de funcionalidade.

Apesar de algumas das plataformas abordadas nos estudos mencionados anteriormente serem as mesmas que iremos avaliar neste trabalho, os critérios avaliados que foram objeto de destaque nestes estudos não coincidem com os mais atuais utilizados pela Gartner em 2017, que serão os utilizados na nossa avaliação.

Adicionalmente, no melhor do nosso conhecimento, a plataforma Knowage nunca foi avaliada com recurso a uma metodologia de avaliação de *software open source*.

Em nenhum dos estudo foi efetuada a instalação das plataformas Knowage e Pentaho, permitindo uma melhor avaliação sobre este processo.

## **4 PLATAFORMAS OPEN SOURCE DE BUSINESS INTELLIGENCE**

Atualmente, encontramos no mercado uma grande diversidade de soluções de BI (comerciais ou *open source*), todas com o objetivo comum de agrupar grandes quantidades de dados, organizando-os de forma clara, transformando-os em informação útil e de qualidade para uma melhor tomada de decisão.

Dentro das plataformas comerciais são exemplo: PowerBI, Tableau, SAP Business Objects, QlikView, IBM Cognos ou Oracle BI.

Relativamente a plataformas *open source* BI, numa primeira fase realizámos uma seleção das que iríamos integrar neste projeto. Para isso, foram considerados seis estudos comparativos que incluíam plataformas *open source* BI (Brandão et al., 2016; T. Ferreira et al., 2017; Lapa et al., 2014; Marinheiro & Bernardino, 2013, 2015; Tereso & Bernardino, 2011). Destes estudos foram selecionadas as seis plataformas mais citadas: Jaspersoft, OpenI, Palo, Pentaho, SpagoBI e Vanilla. Neste conjunto, foram identificadas as plataformas cujos projetos *open source* continuam atualmente ativos e em desenvolvimento bem como as que foram descontinuadas. Assim, as plataformas selecionadas para serem comparadas foram: Jaspersoft, Knowage (SpagoBI) e Pentaho. Apenas estas prosseguem com o desenvolvimento de versões *open source*.

Seguidamente iremos abordar as principais características e funcionalidades das mesmas.

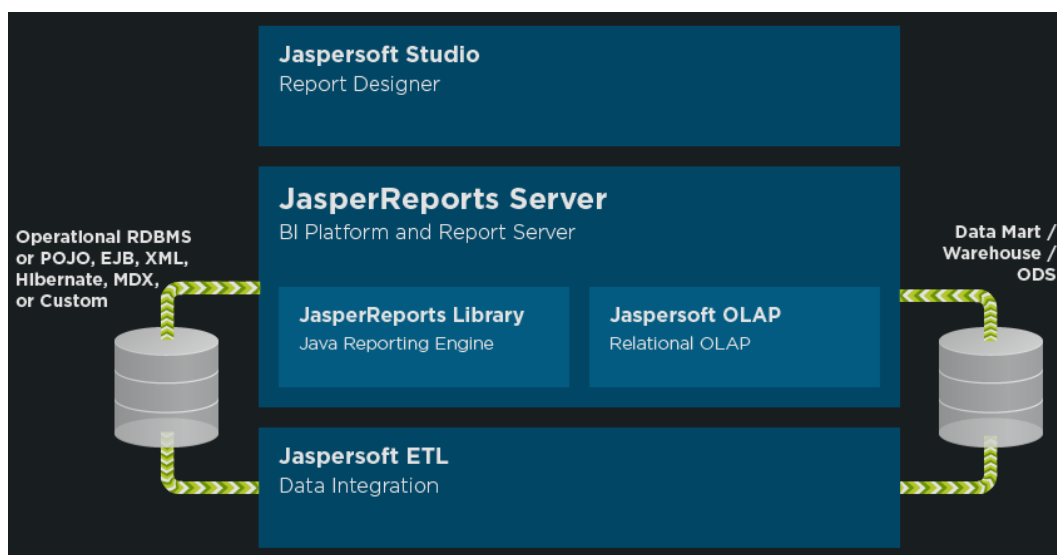
### **4.1 Jaspersoft**

A Jaspersoft foi criada em 2004 com as ferramentas *Jasper Reports* e *iReport Designer* (Jaspersoft, 2018c). Posteriormente foram adicionadas a *JasperReport Server* e a *Jaspersoft OLAP* que em conjunto com as anteriores dão origem, em 2006, à *Jaspersoft Business Intelligence Suite*. Em 2014 a Jaspersoft foi adquirida pela TIBCO Software.

Esta plataforma é totalmente desenvolvida em Java e suporta Windows, Linux, Mac OS X, Apple iOS e Android. Conta atualmente com cinco versões: *Community*, *Reporting*, *AWS (Amazon Web Services)*, *Professional* e *Enterprise*. A TIBCO Jaspersoft assume um modelo de negócio *Open Source* comercial, sendo que das cinco versões apenas a *Community* é disponibilizada em licença *Open Source* (AGPL).

A Figura 4.1 mostra a arquitetura da versão *Community* da Jaspersoft. Conforme ilustrado na figura, a plataforma disponibiliza atualmente cinco componentes: *JaspersoftReports*

*Server, JasperReports Library, Jaspersoft Studio, iReport Designer, Jaspersoft OLAP e Jaspersoft ETL.*



**Figura 4.1** Arquitetura da Plataforma Jaspersoft  
**Fonte:** (Jaspersoft, 2018a)

A *Jaspersoft iReport Designer* permite a elaboração de relatórios para o *JasperReports Server*. Permite a criação de relatórios completos, com gráficos, imagens e tabelas que podem ser exportados para um alargado conjunto de formatos que inclui PDF, XML, HTML, bem como para formatos MS Office e OpenOffice. A *JaspersoftStudio* contém as mesmas funcionalidades que o *Jaspersoft iReport Designer* diferindo que é baseado no *Eclipse*. A *JasperReports Library* é um motor de geração de relatórios *open source* que consegue utilizar informação de qualquer fonte e produzir documentos que podem ser visualizados e exportados numa variedade de formatos. A *JasperSoftReports Server* é um servidor de relatórios que permite o acesso a relatórios e análises que podem ser incorporadas numa página *web* ou aplicação móvel, disponibilizando informação em tempo real ou agendada para acesso em várias plataformas e formatos. Está desenhado por forma a partilhar, proteger e gerir centralmente os seus relatórios.

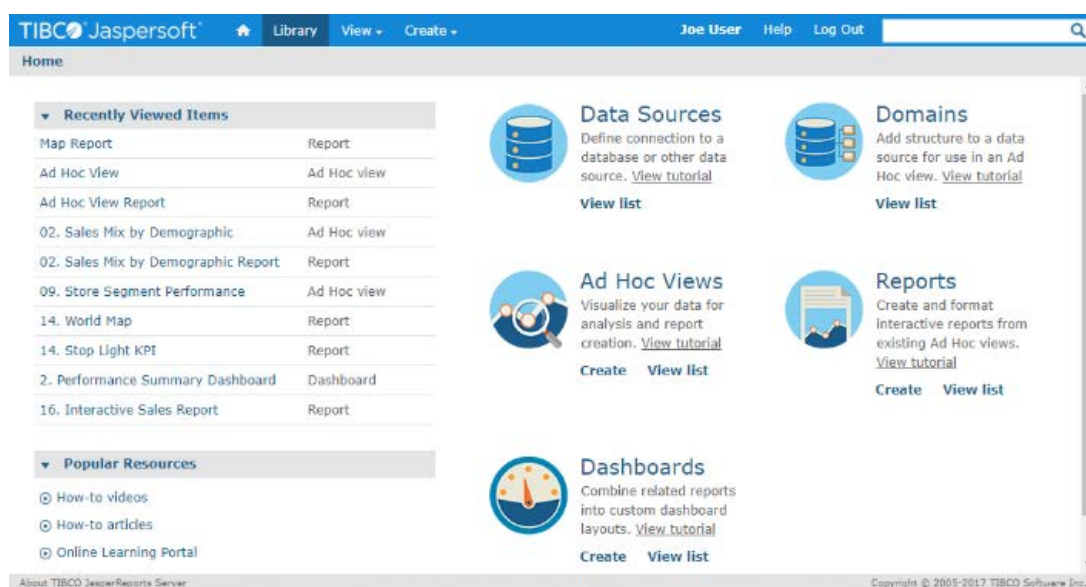
A *Jaspersoft OLAP* consiste num motor OLAP que disponibiliza um ambiente interativo aos utilizadores, onde podem efetuar operações de *slice and dice*, *pivot*, filtragem e resumos de informação em tempo real, através de uma interface *web* ou MS Excel. A *Jaspersoft ETL* é utilizada para extrair informação dos sistemas transacionais e criar *Data Warehouses* e *Data Marts*, posteriormente utilizados para relatórios e análises (Brandão et al., 2016).

O formato modular da instalação permite uma construção da plataforma “à medida” do projeto de BI, o que pode ser visto como uma vantagem, por um lado, ou como uma desvantagem para quem pretende instalar todos os módulos, já que face a soluções de instalação única, a instalação é mais demorada. Esta ferramenta (na versão *Community*) apresenta três lacunas: ausência de um módulo de *Data Mining*, funcionalidades de KPIs e *Dashboards*.

Esta versão está ainda limitada face às versões comerciais na obtenção de análises *In-Memory* e embora permita interatividade nos relatórios, não é possível guardar o resultado gerado com essa interação (*write back*).

A documentação de suporte é extensa e completa, tendo esta ferramenta uma comunidade ativa.

A Figura 4.2 mostra um exemplo da interface da *JasperReports Server*.



**Figura 4.2** Interface JaspersoftReports Server

**Fonte:** (Jaspersoft, 2018b)

## 4.2 Knowage (SpagoBI)

Em 2004, a SpagoWorld, uma iniciativa *open source*, fundada pela Engineering Group (Engineering, 2018), desenvolveu, em *Java*, a plataforma SpagoBI. Desde junho de 2017, aquando do lançamento da versão 6.0, que a plataforma SpagoBI assumiu a denominação de Knowage (Knowage, 2018a). A partir desse momento passou também a existir uma versão comercial (*Enterprise Edition*) e uma versão *open source* (*Community Edition*) sob licença AGPL v3 deixando de ser 100% *open source*.



A Knowage *Community Edition* (CE) mantém todas as funcionalidades da SpagoBI: Reports, OLAP, Gráficos, KPIs, *Dashboards* interativos, GEO/GIS, *Data Mining*, ETL, integração com MS Office e integração *mobile*.

A plataforma é composta pelos módulos *Big Data*, *Smart Intelligence*, *Enterprise Reporting* (ver Figura 4.3), *Location Intelligence*, *Performance Management* e *Predictive Analysis* que segundo a Knowage, permitem melhor escalabilidade. O módulo *Big Data* permite não só trabalhar com grandes volumes de dados, mas também combinar diferentes fontes por modo a desenvolver diferentes análises. O *Smart Intelligence* permite o desenvolvimento de relatórios estáticos, mapas, *cockpits* interativos bem como *queries ad-hoc* via *drag & drop* e análises multidimensionais (OLAP). A versão CE não permite, no entanto, campos calculados, séries temporais e funções MDX (*Multidimensional Expressions*) ao nível de funcionalidades OLAP. O módulo *Enterprise Reporting* faz a apresentação dos relatórios e permite a exportação para diversos formatos incluindo PDF e MS Office. Permite também agendar a emissão de relatórios em modo off-line e distribuir os mesmos pelos utilizadores que se pretenda. A *Location Intelligence* é um módulo dedicado à análise espacial de informação, utilizando vários tipos de bases como mapas ou imagens vetoriais (SVG). Permite trabalhar informação tradicional com informação espacial que tenham uma relação entre si, produzindo mapas dinâmicos.

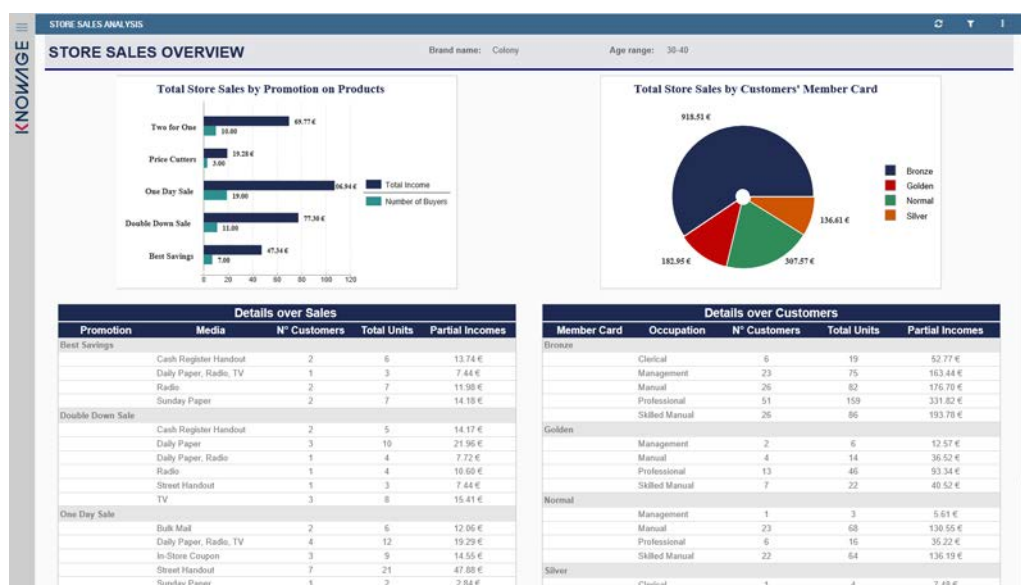


Figura 4.3 Exemplo de relatório da plataforma Knowage  
Fonte: (Knowage, 2018b)

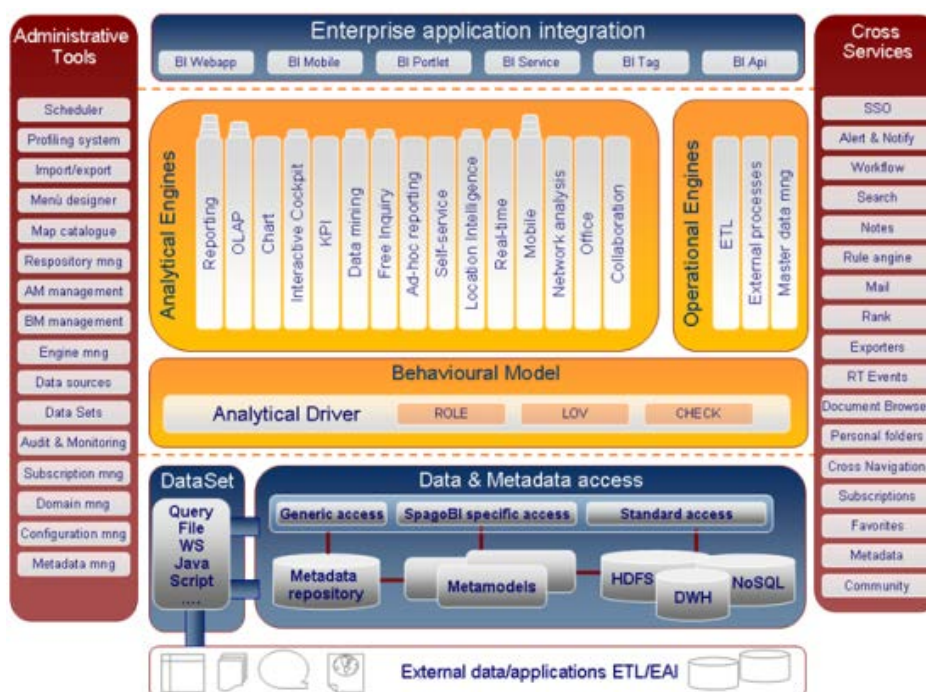
O *Performance Management* é um módulo dedicado à produção e visualização de KPIs e *scorecards*. A *Predictive Analysis* permite um processamento avançado com técnicas

de *Data Mining* para simular ações e avaliar os seus efeitos. Para o efeito *what-if*, este módulo recorre a uma solução OLAP que permite a simulação interativa entre medidas e dimensões via *drag & drop*.

A versão comercial contém os mesmos módulos e funcionalidades que a versão *open source*, acrescentando funções avançadas em quase todas as funcionalidades. São exemplo disso, gráficos mais interativos nos quais é possível efetuar zoom, *cockpits* com atualização quase em tempo real, *what-if* com acesso a MOLAP, *self-service* KPIs. Ao nível da administração da ferramenta, a versão comercial permite instalação multi-ambiente, gestor de *cache* e gestão multi-pessoal.

A plataforma Knowage apresenta-se como uma solução de instalação única. Apenas com uma instalação no servidor, a plataforma fica pronta a operar através do *browser*. Além da versão única, a Knowage disponibiliza os módulos de forma independente, o que a torna bastante versátil no processo de instalação.

A Figura 4.4 mostra a arquitetura da plataforma Knowage (SpagoBI).



**Figura 4.4** Arquitetura da plataforma Knowage (SpagoBI)  
**Fonte:** (SpagoBI, 2014)

A *Community Edition* está bastante completa e a instalação *All-in-One*, acompanhada de um manual extenso e completo, é um ponto forte desta plataforma.

### 4.3 Pentaho

A Pentaho foi criada em 2004, sendo composta pelas ferramentas *Pentaho Reporting*, *Pentaho Reporting Server*, *Mondrian OLAP Server* e *Pentaho Data Integration* (Pentaho, 2018b). Estas ferramentas constituíam a Pentaho Open BI Suite. Em 2006 a Pentaho engloba os projetos *Kettle* e *Weka*. Em 2015, a Pentaho foi adquirida pela Hitachi Data Systems. Nos últimos anos tem sido lançada uma nova edição por ano, estando atualmente na versão 8.0.

A plataforma de BI da Pentaho está disponível em duas versões, ambas desenvolvidas em Java. A *Enterprise Edition*, sendo esta a comercial e a *Community Edition*, a versão *open source*. A plataforma integra os módulos *Business Analytics Platform*, *Data Integration*, *Report Designer* (ver Figura 4.5), *Aggregation Designer*, *Schema Workbench* e *Metadata Editor*. O módulo de *Business Analytic Platform* providencia o servidor que disponibiliza diversos serviços aos utilizadores tais como relatórios e ferramentas de integração. A *Data Integration* constitui o módulo ETL da plataforma, também conhecido como *Kettle* e permite ações de extração, transformação e carregamento de dados. O *Report Designer* é uma ferramenta gráfica que permite conceber relatórios. O *Aggregation Designer* permite criar e manter tabelas agregadas. A *Schema Workbench* é uma interface visual para criar e testar cubos OLAP em Mondrian. E o *Metadata Editor* apresenta-se como uma ferramenta que simplifica a experiência na criação de relatórios permitindo a construção de domínio de *metadata* e modelos relacionais de dados.

Vendor	Bar Code	SKU	Name	Scale	On Hand	Cost	MSRP
AUTOART STUDIO DESIGN		S12_1099	1968 Ford Mustang	1:12	68 units	\$ 95	\$ 195
<b>Description:</b> Hood, doors and trunk all open to reveal highly detailed interior features. Steering wheel actually turns the front wheels. Color dark green.							
					68 units		
CAROUSEL DIECAST LEGENDS		S24_1628	1966 Shelby Cobra 427 S/C	1:24	8.197 units	\$ 29	\$ 50
<b>Description:</b> This diecast model of the 1966 Shelby Cobra 427 S/C includes many authentic details and operating parts. The 1:24 scale model of this iconic lightweight sports car from the 1960s comes in silver and it's own display case.							
					8.197 units		
		S24_2840	1958 Chevy Corvette Limited Edition	1:24	2.542 units	\$ 16	\$ 35
<b>Description:</b> The operating parts of this 1958 Chevy Corvette Limited Edition are particularly delicate due to their precise scale and require special care and attention. Features rotating wheels, working steering, opening doors and trunk. Color dark green.							
					2.542 units		

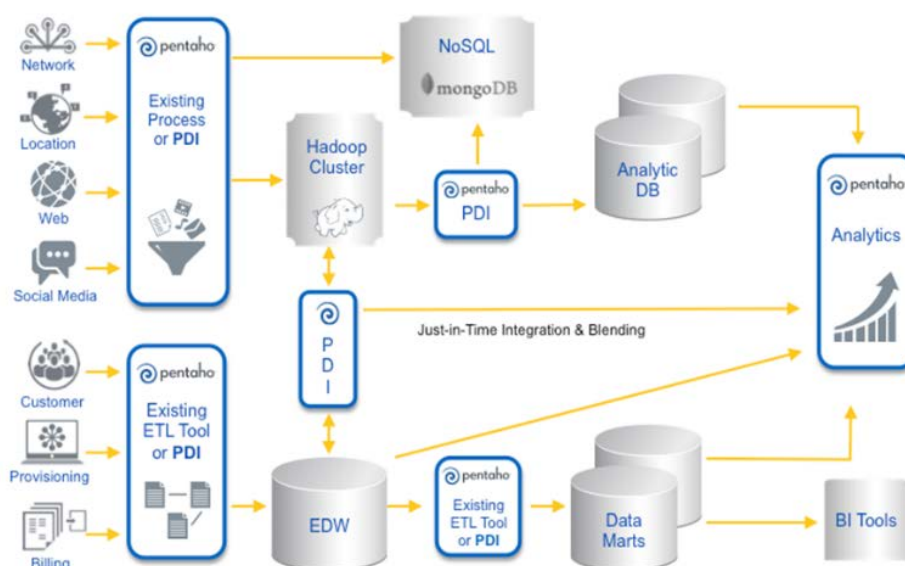
**Figura 4.5** Exemplo de relatório da plataforma Pentaho

Fonte: (Pentaho, 2018a)

A Pentaho destaca funcionalidades que estão presentes apenas na versão comercial. Entre elas, temos os relatórios interativos, *ad-hoc queries*, *Drill down* e *Drill through*, GEO/GIS, *Dashboards* e aplicação *mobile*. Destacam ainda opções mais avançadas na integração de dados e mais aplicações em *Big Data*.

É, no entanto, possível a implementação de *Dashboards* com recurso às *Community Tools* e efetuar *Data Mining* instalando o módulo *Weka*.

A Figura 4.6 mostra a arquitetura da plataforma Pentaho.



**Figura 4.6** Arquitetura da plataforma Pentaho  
**Fonte:**(Pentaho, 2018a)

Tal como a Jaspersoft, o formato modular da arquitetura e instalação permite uma construção da ferramenta plataforma “à medida” do projeto, o que pode ser visto como uma vantagem por um lado, ou como uma desvantagem para quem pretende instalar todos os módulos já que face a soluções de instalação única é mais demorada.

A documentação de suporte é extensa, incluindo uma plataforma *wiki*, e uma comunidade de utilizadores bastante ativa.



## **5 METODOLOGIAS DE AVALIAÇÃO DE SOFTWARE OPEN SOURCE**

Neste capítulo abordamos as diferentes metodologias de avaliação de software open source existentes. Procedemos inicialmente à escolha de uma para utilização posterior na avaliação das plataformas open source de BI e de seguida à sua descrição.

### **5.1 Escolha de Metodologia**

Uma das principais preocupações da Engenharia de *Software* é a produção de *software* de qualidade, portanto, a avaliação dessa qualidade sempre foi uma tarefa crítica para os profissionais desta área. Essas avaliações são baseadas em modelos de *software* que definem e medem a qualidade do *software*, combinando métricas de *software* com opiniões de especialistas (Samoladas & Gousios, 2008).

O advento do FLOSS fez com que os modelos tradicionais de avaliação de *software* como McCall, Boehm ou ISO 9126 não fossem aplicáveis a todo o *software*. Estes modelos não podem ser adaptados às práticas de desenvolvimento de *software open source* e, portanto, não podem ser usados para avaliar o *software* e a sua comunidade como um todo (Samoladas & Gousios, 2008).

Samoladas e Gousios (2008) realizaram um estudo onde identificaram quatro metodologias para avaliação de *software open source*: *Software Quality Observatory for Open Source Software* (SQO-OSS), *Open Source Maturity Model* (OSMM), *Open Business Readiness Rating* (OpenBRR) e *Qualification and Selection of Open Source Software* (QSOS). Baseadas nas metodologias mencionadas, surgiram posteriormente duas novas: *Open Business Quality Rating* (Open BQR) e OSSpal.

Das seis metodologias referidas, a SQO-OSS, a OSMM e a Open BQR, além de serem metodologias antigas, são atualmente pouco utilizadas ou não utilizadas de todo. Pouca informação é encontrada sobre estas metodologias e os *websites* oficiais já não estão ativos.

Segundo Samoladas & Gousios (2008), a QSOS retorna resultados objetivos, mas não é flexível e é difícil de operar. Embora existam alguns trabalhos recentes utilizando a QSOS (Ferreira, Ferros, & Fernandes, 2012; Nereu, 2017), esta metodologia tem agora mais de uma década sem qualquer atualização. Por outro lado, a OSSpal surge em 2014 como uma revisão da OpenBRR, mantendo o objetivo de ajudar as empresas a encontrar

*software open source* de alta qualidade (Wasserman, Guo, Mcmillian, & Qian, 2017). Consideramos, portanto, que a OSSpal é a metodologia mais adequada atualmente para a avaliação de *software open source* no contexto do trabalho que pretendemos desenvolver.

## 5.2 OSSpal

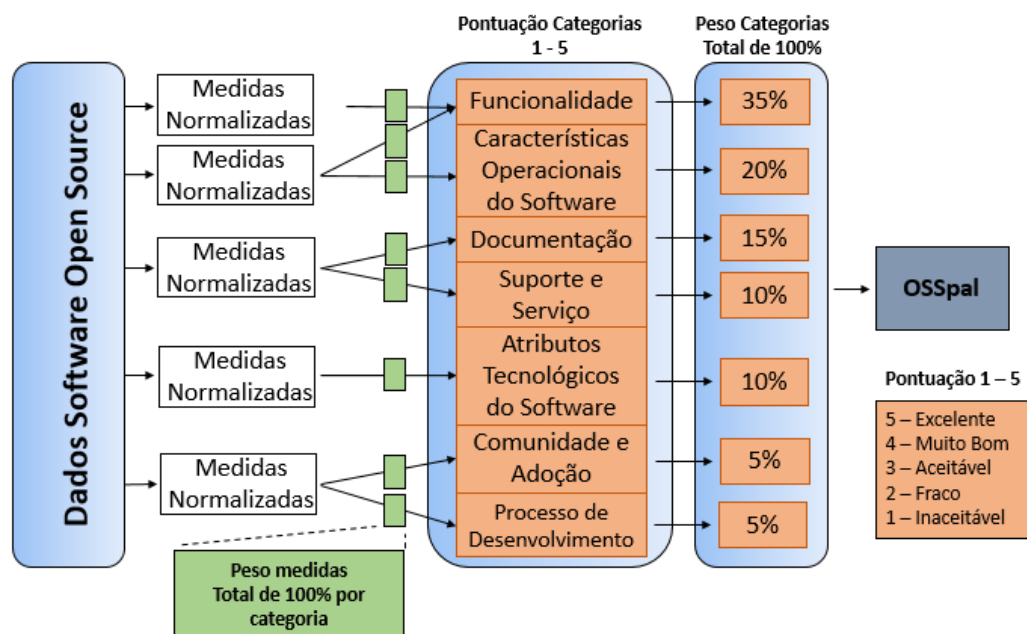
A OSSpal emergiu como um sucessor do *Business Readiness Rating* (OpenBRR) com o objetivo de fornecer uma fonte confiável e imparcial para avaliação de *software open source*. O objetivo é ser um modelo de avaliação aberto, abrangente e padrão que seja confiável, amplamente utilizado e ajustável (Wasserman, 2014).

A OSSpal combina medidas de avaliação quantitativa e qualitativa que nos permitem identificar qual o *software* com melhor pontuação. Desta forma, pode ajudar empresas, agências governamentais e outras organizações a encontrar *software open source* de alta qualidade (Wasserman et al., 2017).

A metodologia OSSpal, apresentada na Figura 5.1, consiste em sete categorias de avaliação (Wasserman et al., 2017):

- **Funcionalidade:** Quão bem o *software* dará resposta às necessidades consideradas essenciais pelos utilizadores?
- **Características Operacionais do Software:** Quão seguro é o *software*? Quão bem o *software* funciona? Quão bem o *software* está dimensionado para um ambiente grande? Quão boa é a interface de utilizador? Quão fácil de usar é o *software* para utilizadores finais? Quão fácil é o *software* para instalar, configurar, implementar e manter?
- **Suporte e serviço:** Quão bem o componente de *software* é suportado? Existe suporte comercial e / ou comunitário? Existem pessoas e organizações que podem fornecer serviços de formação e consultoria?
- **Documentação:** Existe tutorial e documentação de referência adequados para o *software*?
- **Atributos Tecnológicos do Software:** Quão bem está o *software* arquitetado? Quão modular, portátil, flexível, extensível, aberto e fácil de integrar? O design, o código e os testes são de alta qualidade? Quão completos e livres de erros eles são?

- **Comunidade e Adoção:** Quão bem está o componente adotado pela comunidade, mercado e indústria? Quão ativa é a comunidade para o software?
- **Processo de Desenvolvimento:** Qual o nível de profissionalismo do processo de desenvolvimento e da organização do projeto como um todo?



**Figura 5.1** Metodologia OSSpal  
**Fonte:** Adaptado de Wasserman (2014)

A implementação da Metodologia OSSpal é composta por quatro fases (OpenBRR, 2005; Wasserman, 2014; Wasserman et al., 2017). Na fase 1 procedemos à identificação das características que pretendemos ver avaliadas no *software* em análise. Tendo em conta o tipo de *software*, deve ser criado um conjunto de características específicas para avaliação. Na fase 2 procedemos à atribuição de pesos às categorias e medidas. Deve ser atribuída uma percentagem de acordo com a importância que queremos dar a cada categoria. O total das percentagens deve totalizar 100%. Caso sejam definidas medidas dentro de alguma categoria deve-se proceder do mesmo modo atribuindo uma percentagem de importância a cada medida num total de 100% dentro de cada categoria. Na fase 3 procede-se à recolha de informação que permita atribuir uma ponderação a cada medida ou categoria num nível de 1 (inaceitável) a 5 (excelente). Na fase 4 procede-se ao cálculo da pontuação final da OSSpal aplicando todos os pesos definidos antes.

A Funcionalidade é uma categoria de avaliação que é calculada de maneira diferente das restantes categorias. Cada tipo de *software* possui um conjunto próprio de



características/funcionalidades que necessitam ser preenchidas. A classificação da categoria funcionalidade é obtida comparando as características do *software* avaliado com o conjunto de características padrão necessário para um uso elementar. Esse conjunto de características padrão deve ser construído ou obtido de uma fonte externa (fase 1).

As etapas seguintes devem ser utilizadas para calcular a pontuação da categoria funcionalidade:

- i. Atribuir uma pontuação correspondente ao nível de importância a todos os itens na lista de características, usando uma escala de 1 a 3, sendo 1 menos importante e 3 muito importante.
- ii. Comparar a lista de características do *software* avaliado com a lista de características padrão. Para cada característica validada, incluir a pontuação de importância numa soma cumulativa. Se não for validada, deduzir a pontuação de importância da soma.
- iii. Dividir a soma cumulativa obtida pela soma cumulativa máxima que se obteria pela validação de todas as características considerando o nível máximo. A este rácio chamamos pontuação das características.
- iv. Normalizar a pontuação das características para a escala de 1 a 5 utilizando o seguinte esquema:
  - Superior a 96%, pontuação = 5 (excelente)
  - 90% - 96%, pontuação = 4 (muito bom)
  - 80% - 90%, pontuação = 3 (aceitável)
  - 65% - 80%, pontuação = 2 (fraco)
  - Menos de 65%, pontuação = 1 (inaceitável)

Este processo é executado sobre cada *software* em avaliação, devolvendo a pontuação da categoria funcionalidade de cada um.

## **6 SELEÇÃO DA PLATAFORMA OPEN SOURCE DE BI**

A primeira parte deste capítulo será dedicada à comparação das três plataformas *open source* BI descritas no capítulo 4, tendo em conta as funcionalidades das plataformas. Posteriormente, e tendo em conta os resultados desta comparação, iremos realizar uma avaliação das duas melhores plataformas, utilizando a metodologia de avaliação de *software open source* descrita no capítulo 5, a OSSpal. Após esta avaliação, a plataforma que se apresentar como melhor solução, será então a escolhida para ser implementada.

### **6.1 Comparação das plataformas**

No processo de comparação entre plataformas de BI, foram considerados os critérios definidos pela Gartner no Quadrante Mágico 2017 (Sallam et al., 2017). Os critérios da Gartner de 2013 foram revistos em 2016, tendo em conta mais de uma década de evolução na área do *Business Intelligence* e, no melhor do nosso conhecimento, esta é a primeira avaliação de plataformas *open source* de BI segundo os novos critérios.

A Gartner cataloga estes critérios em 5 categorias: Infraestrutura, Gestão de Dados, Criação de Análises e Conteúdo, Partilha de Informação e Capacidades Gerais, num total de 15 critérios. Desses, foram selecionados 11: *Cloud* BI, OLAP, Fontes *Hadoop/NoSQL*, ETL, *Self-Service* BI, Informação em Tempo Real, *Dashboards*, Visualização Interativa, *Mobile* BI, Colaboração e Instalação “*All-in-One*”, uma vez que são os que apresentam maior objetividade e facilidade de avaliação. Foram excluídos 4 critérios: Administração e Segurança, Gestão de Metadados, *Smart Data Discovery* e Facilidade de Uso, devido à sua avaliação ter um carácter mais subjetivo.

Os critérios selecionados podem ser definidos da seguinte forma:

- **Cloud BI** – A plataforma de *Business Intelligence* é acedida através de um *browser*, estando alojada em servidores locais ou externos (*Internet*).
- **OLAP** – Permite que os utilizadores analisem dados com pesquisas bastante rápidas, multidimensionais, disponibilizando um estilo de análise conhecido como “*slicing and dicing*”.
- **Fontes Hadoop/NoSQL** – Capacidade para se ligar a fontes *Hadoop* e *NoSQL* (*BigData*).

- **ETL** – Permite efetuar a extração de dados de diversos sistemas e fontes, transformar esses dados segundo os modelos definidos, e efetuar o seu carregamento para *Data Marts* ou *Data Warehouses*.
- **Self-Service BI** – Permite que os utilizadores façam as suas perguntas de dados, sem necessitar de um profissional de Tecnologias da Informação para criar um relatório.
- **Informação em Tempo Real** – Capacidade de devolver informação processada em tempo real nos diversos meios (*Dashboards, reports, etc.*)
- **Dashboards** – Capacidade de publicar, baseado em *Web* ou dispositivos móveis, relatórios com mostradores interativos de informação. Estes mostradores indicam o estado das métricas de desempenho comparadas com um objetivo.
- **Visualização interativa** – Oferece ao utilizador a possibilidade de visualizar numerosas vistas dos dados mais eficientemente, usando imagens e gráficos, em vez de linhas e colunas.
- **Mobile BI** – Esta funcionalidade permite que as organizações apresentem relatórios e *Dashboards* nas plataformas móveis, numa publicação e/ou modo interativo, aproveitando o modo de interação dos dispositivos e outras capacidades não disponíveis normalmente nos monitores.
- **Colaboração** – Permite que os utilizadores de BI partilhem e discutam a informação disponibilizada, através de *chats* e notas, e que a possam integrar noutras ferramentas sociais ou colaborativas.
- **Instalação “All-in-One”** – Disponibiliza a instalação completa da plataforma através de um único ato de instalação por comparação a um processo de instalação por módulos.

Cada plataforma, na sua versão *open source*, foi avaliada em cada critério. A avaliação foi efetuada com base na informação disponibilizada por cada plataforma no seu *website* e comunidades *open source*.

A Tabela 6.1 resume a comparação das plataformas, segundo os critérios definidos, assinalando-se a evidência de cada um dos critérios associados às plataformas consideradas (usando o símbolo ✓).

**Tabela 6.1** Comparação Plataformas *Open Source* BI

Critérios	Plataformas <i>Open Source</i> de BI		
	<i>Jaspersoft</i>	<i>Knowage</i>	<i>Pentaho</i>
<i>Cloud</i> BI	✓	✓	✓
OLAP	✓	✓	✓
<i>Hadoop/NoSQL</i>		✓	✓
ETL	✓	✓	✓
<i>Self-Service</i> BI		✓	
Informação em Tempo Real	✓	✓	✓
<i>Dashboards</i>	✓	✓	✓
Visualização Interativa		✓	
<i>Mobile</i> BI	✓	✓	
Colaboração			
Instalação “ <i>All-in-One</i> ”		✓	
<b>Total</b>	<b>6</b>	<b>10</b>	<b>6</b>

Desta comparação destaca-se, pelo número de critérios que cumpre, a plataforma Knowage, que valida 10 dos 11 critérios. A Jaspersoft e a Pentaho cumprem, 6 dos 11 critérios. Atendendo a que são praticamente os mesmos, pode considerar-se que as duas plataformas são muito equivalentes nesta avaliação.

As três plataformas cumprem, em simultâneo, os critérios *Cloud* BI, OLAP, ETL, Informação em Tempo Real e *Dashboards*. Estes são critérios sobre funcionalidades que, não sendo recentes, são relevantes para a estrutura de uma plataforma de BI e é, por isso, natural a sua presença nas três plataformas. Ainda assim, a obtenção de *Dashboards* na Pentaho só é possível de forma livre com recurso às *Community Tools* desenvolvidas pela comunidade *open source* da Pentaho. Em *Mobile* BI, só a Pentaho tem esta funcionalidade reservada à plataforma comercial.

No âmbito da ligação a fontes *BigData*, no critério *Hadoop/NoSQL*, a Jaspersoft só disponibiliza este tipo de ligação na versão comercial da plataforma. Sobre a disponibilidade de *Self-Service* BI, a Knowage destaca-se, sendo a única a disponibilizar

em *open source* esta funcionalidade. Tendo em conta a importância atual nas organizações do *Self-Service BI*, este é um ponto forte a registar na Knowage. Do mesmo modo, enquanto a Jaspersoft e a Pentaho só disponibilizam visualizações interativas na versão comercial, a Knowage disponibiliza visualizações interativas mais avançadas na sua versão *open source*.

Pela definição de colaboração da Gartner, nenhuma das plataformas verifica este critério, sendo que apenas a Knowage apresenta referências à colaboração na sua versão comercial.

Um aspeto importante das plataformas é o processo de instalação onde apenas a Knowage apresenta a possibilidade de instalação “*All-in-One*”, sendo este um ponto forte. Esta funcionalidade reduz a complexidade do processo e também a possibilidade de falhas na integração entre módulos.

Com a aplicação dos novos critérios da Gartner ficam visíveis diferenças que, aplicando critérios antigos, não seriam detetáveis. No que diz respeito a novas tecnologias e funcionalidades, a Jaspersoft e a Pentaho relegam-nas em exclusivo para a sua versão comercial, sendo por isso as plataformas com a maior diferença entre as versões comercial e *open source*. No lado oposto temos a Knowage que apresenta uma versão *open source* muito idêntica à versão comercial.

Das três plataformas comparadas, a Knowage é a mais completa, destacando-se por disponibilizar uma versão de instalação “*All-in-One*” que agiliza todo o processo.

## **6.2 Avaliação das plataformas**

Após comparadas as plataformas *open source* de BI, tendo em conta os critérios da Gartner 2017 e a informação disponibilizada por cada fornecedor, iremos agora avaliar as duas que validam mais critérios, utilizando uma metodologia para avaliação de *software open source*: a metodologia OSSpal.

A Knowage obteve o melhor resultado na comparação efetuada, enquanto a Jaspersoft e a Pentaho tiveram um desempenho idêntico. Com base nesses resultados, a Knowage é a primeira plataforma de BI selecionada para a avaliação. Em vários estudos recentes (T. Ferreira et al., 2017; Marinheiro & Bernardino, 2015; Tereso & Bernardino, 2011) a Pentaho obtém melhores pontuações do que a Jaspersoft. Além disso, uma vez comparado no Google Trends, o Pentaho obtém um resultado de 83 enquanto o Jaspersoft obtém

apenas 20. Assim, o Pentaho é a segunda plataforma de BI selecionada para esta avaliação.

Para garantir uma avaliação mais confiável e precisa usando a OSSpal, procedemos à instalação de ambas as plataformas. A instalação foi seguida por um uso elementar por forma a obter experiência de utilizador.

Conforme indicado na fase 1 da metodologia OSSpal, foi elaborada uma lista de características para a categoria funcionalidade. Selecionámos a nossa lista de características seguindo os mesmos critérios usados na comparação anterior. Com a lista de características elaborada, foi atribuído um peso de 1 (pouco importante) a 3 (muito importante) a cada característica.

A Tabela 6.2 mostra as características escolhidas para a categoria de Funcionalidade e os pesos atribuídos a cada uma, de acordo com a metodologia OSSpal.

**Tabela 6.2** Pesos atribuídos a cada característica na categoria Funcionalidade

<b>Característica</b>	<b>Peso</b>
<i>Dashboards</i>	3
Visualização Interativa	3
OLAP	3
Informação em Tempo Real	3
ETL	2
<i>Mobile BI</i>	2
<i>Self-Service BI</i>	2
Instalação “ <i>All-in-One</i> ”	1
<i>Cloud BI</i>	1
Colaboração	1
<i>Hadoop/NoSQL</i>	1

Conforme indicado na fase 2, alocámos pesos a cada categoria totalizando 100%, conforme mostrado na Tabela 6.3.

Considerámos “Funcionalidade” a categoria mais importante, pois consiste no núcleo do *software* avaliado. Por este motivo, foi dado o maior peso (35%). Seguindo com um peso de 20%, consideramos “Características Operacionais do *Software*” a segunda categoria mais importante pois engloba áreas para avaliação como a experiência de utilizador e o processo de instalação.

**Tabela 6.3** Pesos atribuídos a cada categoria

<b>Categoria</b>	<b>Peso</b>
Funcionalidade	35%
Caraterísticas Operacionais do <i>Software</i>	20%
Documentação	15%
Suporte e Serviço	10%
Atributos Tecnológicos do <i>Software</i>	10%
Comunidade e Adoção	5%
Processo de Desenvolvimento	5%
<b>Total</b>	<b>100%</b>

Ainda com alguma importância, com pesos de 15% e 10%, seguem-se as categorias “Documentação” e “Suporte e Serviço” respetivamente. Especialmente no contexto de *software* livre, essas categorias desempenham um papel importante no apoio aos utilizadores e profissionais de Tecnologias da Informação.

Considerada de menor relevância, a categoria “Atributos Tecnológicos do *Software*” recebeu uma ponderação de 10% e as categorias “Comunidade e Adoção” e “Processo de Desenvolvimento”, ambas receberam um peso de 5%.

Após a atribuição de pesos a todas as categorias, seguimos para a fase 3. Cada plataforma de BI é avaliada e para cada categoria (exceto a funcionalidade, que é processada de forma diferente) é dada uma pontuação de 1 (inaceitável) a 5 (excelente).

Como mencionado antes, a pontuação de 1 a 5 para a categoria funcionalidade é calculada de forma diferente. A Tabela 6.4 apresenta os resultados intermédios deste passo e a pontuação final da categoria funcionalidade.

Na fase 4, todas as pontuações são traduzidas de acordo com o peso que cada categoria recebeu (por exemplo, 10% de 5 traduz-se em 0.5). A soma acumulada de cada pontuação traduzida por categoria fornece a pontuação final de cada plataforma de BI.

A Tabela 6.5 apresenta os resultados da avaliação.

A plataforma Pentaho, com pontuação final de 4.35 (de 1 a 5) foi a plataforma de BI com maior pontuação. No caso da Knowage a pontuação final foi de 3.31.

A pontuação da plataforma Pentaho foi ligeiramente superior à da plataforma Knowage em todas as categorias avaliadas, exceto na categoria Funcionalidade onde foi superior em 0.35 pontos. No primeiro passo para o cálculo da pontuação da categoria

Funcionalidade, a Knowage obteve um resultado de 86% e a Pentaho de 91%. Apesar dos resultados serem bastante próximos (86% e 91%), a normalização para valores entre 1 e 5, aplicada pela metodologia OSSpal, transforma estes valores em 3 para a Knowage e 4 para a Pentaho. Aplicando o peso de 35% atribuído à categoria funcionalidade, resulta na diferença final de 0.35.

**Tabela 6.4** Pontuação na categoria funcionalidade

<i>Característica</i>	<i>Peso</i>	<b>Knowage</b>	<b>Pentaho</b>
<i>Dashboards</i>	3	3	3
Visualização Interativa	3	3	3
OLAP	3	3	3
Informação em Tempo Real	3	3	3
ETL	2	0	2
<i>Mobile BI</i>	2	2	2
<i>Self-Service BI</i>	2	2	2
Instalação “ <i>All-in-One</i> ”	1	1	0
<i>Cloud BI</i>	1	1	1
Colaboração	1	0	0
Hadoop/NoSQL	1	1	1
<b><i>Soma Cumulativa</i></b>	<b>22</b>	<b>19</b>	<b>20</b>
<b><i>Normalização para escala 1 a 5</i></b>	<b>100%</b>	<b>86%</b>	<b>91%</b>
		<b>3</b>	<b>4</b>

A aplicação do peso de 35% a estas pontuações resulta num impacto mais relevante na pontuação final do que na realidade era antes da normalização.

Nas características operacionais do *software*, a Pentaho apresenta uma interface mais simples que a Knowage, no entanto, mais intuitivo e eficaz. Em termos de documentação a Pentaho possui mais e melhor documentação de suporte, o que no contexto do *open source* é bastante importante.

A diferença final entre as pontuações da Pentaho e Knowage é de 1.04. Atribuímos essa diferença ao facto de que a Pentaho tem uma adoção mundial muito maior, o que ajuda a tornar-se num *software* mais “maduro”.

A Pentaho apresentou a melhor pontuação após aplicar a metodologia OSSpal. A Knowage obteve uma pontuação inferior à Pentaho, mas tem todo o potencial de para apresentar um desempenho melhor no futuro.



A Knowage tem um pacote “*All-in-One*” para instalação, o que simplifica o processo e, pela nossa experiência, o processo de instalação e parametrização demora cerca de meia hora. A Pentaho tem mais etapas para atingir o mesmo estágio, mas, se todas as instruções forem seguidas corretamente, da nossa experiência podemos afirmar que pode estar em funcionamento em menos de uma hora.

**Tabela 6.5** Pontuação final da OSSpal

Categoria	Pontuação	
	Knowage	Pentaho
Funcionalidade	1.05	1.40
Características operacionais do <i>software</i>	0.80	1.00
Documentação	0.53	0.68
Suporte e Serviço	0.35	0.45
Atributos Tecnológicos do <i>Software</i>	0.30	0.40
Comunidade e Adoção	0.13	0.23
Processo de desenvolvimento	0.15	0.20
<b><i>TOTAL</i></b>	<b><i>3.31</i></b>	<b><i>4.35</i></b>

A conclusão geral é que a Pentaho é um *software* mais maduro que a Knowage em todas as categorias e isso é o resultado de um uso e comunidade muito maior em todo o mundo. Por este motivo, a Pentaho é a plataforma selecionada para o desenvolvimento do estudo de caso.

## **7 ESTUDO DE CASO**

Neste capítulo iremos desenvolver um estudo de caso em *Business Intelligence* utilizando dados e informação de uma empresa portuguesa. Para o efeito, iremos utilizar o conhecimento adquirido em *Business Intelligence* e a plataforma Pentaho como suporte tecnológico.

Para o levantamento de requisitos e caracterização do negócio foram conduzidas entrevistas a colaboradores da empresa em abril de 2018.

### **7.1 Preservação de anonimato**

Os dados utilizados para integração na plataforma *open source* de BI foram gentilmente cedidos pela empresa. O parecer positivo à colaboração com este projeto foi obtido sob a condição de total anonimato, sendo a empresa designada neste documento por “Good Mailman”, nome fictício. Os termos utilizados neste estudo não são os utilizados no seio da empresa nem os procedimentos são exatamente os reais, servindo para caracterizar o ramo de negócio e a operação da empresa.

Os dados apresentados, sob qualquer forma, foram submetidos a uma ponderação, mantendo a relevância para o estudo sem revelar os dados reais do negócio da Good Mailman.

### **7.2 A empresa**

A Good Mailman (GM) desenvolve a sua atividade na Distribuição Postal Empresarial, tendo os seus principais clientes na área da banca, telecomunicações e águas. A distribuição postal da GM, vulgo, correio postal, opera no sentido *Business to Consumer* (B2C), isto é, nas comunicações via postal das empresas para os seus clientes.

A GM desenvolve a sua atividade no eixo litoral do país, tendo diversos centros localizados estrategicamente ao longo deste eixo.

### **7.3 Contexto Operacional**

A empresas clientes da GM fornecem os dados a uma empresa externa que os processa, imprime, efetua a “envelopagem” e separação por código postal. A GM recolhe as cartas nessa empresa e procede à separação e envio para cada centro. Por sua vez, em cada centro, cada carteiro recolhe as cartas da sua zona para entrega ao cliente final (destinatário da carta).

A divisão territorial é definida com base no código postal de 4 dígitos (ex.: 3030 – Coimbra) designado por CP4. A cada Centro é atribuído um grupo de códigos postais.

A cada carteiro é atribuída uma zona. Esta zona é definida com base no código postal de 7 dígitos (ex.: 3030-853 Rua da Beira, Ceira) designado por CP7. Cada zona é constituída por um conjunto de CP7s.

Aquando do levantamento de um lote de cartas, a empresa que efetuou a sua preparação fornece a informação relativa a esse lote que é importada para o sistema de informação da GM. Cada carta fica com um registo na base de dados.

O carteiro está munido de meios que lhe permitem reportar o estado de cada carta e indicar, por exemplo, o momento em que é entregue.

O sistema de informação da GM assenta atualmente em soluções com licenças proprietárias e/ou desenvolvidas à medida. Dependendo da função de cada colaborador, este tem acesso a diferentes níveis de informação, sobretudo operacional.

Este sistema pode ser acedido através de vários meios e pode difundir informação de forma periódica, por email.

#### **7.4 Análise de requisitos**

Nas entrevistas realizadas a colaboradores da GM, foram identificadas as análises efetuadas/consultadas com mais frequência, sendo estas por: Data, Cliente, Centro, Código Postal, Estado, Idade, Carteiro e Zona. Para o desenho do modelo dimensional, estas serão as dimensões a considerar. Na dimensão Data, a análise mais frequente é por semana.

Importa neste momento contextualizar duas destas dimensões: Estado e Idade.

- **Estado** refere-se à situação em que se encontra a carta no seu ciclo desde a entrada do registo nos sistemas de informação da GM até à sua entrega ao destinatário ou devolução ao remetente por impossibilidade de entrega ao destinatário. Existe um conjunto determinado de estados que podem ser categorizados como “finais” ou “não finais”. Por motivos de confidencialidade, iremos apenas assumir os estados “OK” e “Não OK” nesta dimensão, representando de forma global os estados finais e não finais. Assim, “OK” significa que a carta foi entregue ao destinatário ou

devolvida ao remetente. “Não OK”, por sua vez, significa que a carta está em distribuição para entrega ou em processo de devolução ao remetente.

- **Idade** representa o número de dias desde a importação do registo da carta no sistema de informação da GM até que a carta é entregue ao destinatário ou, caso não esteja entregue, até ao dia da consulta da informação.

Adicionalmente, as análises da GM são efetuadas também sobre o Tempo Médio de Vida e o Volume. Do ponto de vista de análise multidimensional, Tempo Médio de Vida e Volume são consideradas medidas.

O **Tempo Médio de Vida** representará a média de idades das cartas no conjunto que estiver a ser analisado. Este é o principal indicador utilizado para avaliar o nível de serviço. Para a GM, um nível de serviço aceitável significa ter este indicador em cinco ou menos dias.

O **Volume** representa o total de cartas no conjunto que estiver a ser analisado. Esta medida serve sobretudo para identificar zonas subdimensionadas ou sobredimensionadas e permitir um redimensionamento das mesmas. Serve também para acompanhar a evolução do tráfego gerado por cada cliente e o seu impacto na estrutura.

Com base na análise anterior, é proposta a criação de análises relativamente ao Volume de cartas e Tempo Médio de Vida. Estas análises devem variar nas dimensões disponíveis e de acordo com o colaborador e a informação adequada às suas funções.

Com o objetivo de proporcionar ao coordenador do centro uma visão global do mesmo sobre o desempenho passado e atual, foi proposta a criação de uma análise mensal sobre a dimensão estado. Deverão ser fornecidos ao coordenador os Tempos Médios de Vida e Volumes por Cliente, Código Postal e Carteiro.

Com o objetivo de proporcionar uma visão sobre o desempenho dos Centros a colaboradores com funções de controlo geral, foi proposta a criação de uma análise de evolução semanal. Deverão ser fornecidos os Volumes e Tempos Médios de Vida nas dimensões Cliente, Código Postal e Estado.

Adicionalmente propõe-se a criação de um *dashboard* de acompanhamento global da evolução dos Volumes e Tempos Médios de Vida

## 7.5 Tratamento de dados

Segundo Kimball (2013), uma base de dados relacional com esquema em estrela é uma boa base para a construção de um cubo OLAP. Assim, tendo em conta que as análises multidimensionais seriam sobre um cubo OLAP, o tratamento de dados foi conduzido no sentido da criação de um modelo dimensional em estrela. Descrevemos de seguida os passos até à criação deste modelo.

### 7.5.1 Dados fornecidos pela Good Mailman

A Good Mailman forneceu um registo de importações de cartas de 1 de abril até 17 de maio de 2018 e apenas das zonas de Coimbra e Leiria. O número de registos situa-se nas centenas de milhar e foram facultados em formato CSV. Os campos fornecidos de cada carta foram: data de importação no sistema, data de entrega, código postal da morada de entrega (destinatário) em formato de 7 dígitos, cliente da Good Mailman que efetua o envio (remetente) e o centro responsável pela entrega. A Figura 7.1 ilustra os primeiros registos deste ficheiro.

	A	B	C	D	E
1	data_imp	data_entrega	codigo_postal7	cliente	centro
2	15/05/2018	NULL	3030-183	Cliente A	Coimbra
3	16/05/2018	NULL	3030-177	Cliente A	Coimbra
4	16/05/2018	17/05/2018	2490-132	Cliente A	Leiria
5	27/04/2018	08/05/2018	2415-520	Cliente B	Leiria
6	27/04/2018	08/05/2018	2415-520	Cliente B	Leiria
7	27/04/2018	09/05/2018	2415-576	Cliente B	Leiria
8	27/04/2018	08/05/2018	2415-520	Cliente B	Leiria
9	27/04/2018	08/05/2018	2415-520	Cliente B	Leiria
10	27/04/2018	09/05/2018	2430-520	Cliente B	Leiria
11	01/05/2018	09/05/2018	2460-091	Cliente B	Leiria
12	01/05/2018	09/05/2018	3000-279	Cliente B	Coimbra
13	01/05/2018	09/05/2018	3030-397	Cliente B	Coimbra
14	01/05/2018	09/05/2018	3030-397	Cliente B	Coimbra

**Figura 7.1** Excerto dos registos fornecidos pela Good Mailman

### 7.5.2 Campos adicionais

Perante um número reduzido de campos, foram introduzidos novos campos com dados que se podiam extrair dos existentes. Isto tornou possível aumentar o número de análises disponível sobre um conjunto bastante limitado de dados. Foram criados e/ou revistos os seguintes campos:

- Adicionado campo com um número de ordem sequencial, identificador único para cada carta, `id_carta`.
- No campo `data_entrega` foram removidas todas as ocorrências da string “*NULL*”. A partir deste campo foi criado o campo `status`. Se

existir uma data de entrega, significa que a carta tem um estado final, já classificado por nós como “OK”. Caso não exista, a carta não tem um estado final e é considerada “Não OK”.

- O campo `codigo_postal7` apresenta um detalhe demasiado elevado para as análises que pretendemos efetuar e como tal foi reduzido ao código postal de 4 dígitos.
- Para ser possível a análise de um dos principais indicadores de nível de serviço, foi criado e preenchido o campo `idade`. Se existir uma data de entrega, isto é, de um estado final, a idade, em dias, é obtida pela diferença entre a data de entrega (`data_entrega`) e a data de importação (`data_imp`). Se não existir uma data de entrega, a idade é obtida pela diferença do dia da consulta da análise e a data de importação. Uma vez que os dados são estáticos e representam a realidade a dia 17 de maio de 2018, no caso de inexistência de data de entrega, a idade foi calculada pela diferença entre a data de importação e 17 de maio de 2018.

Após estes passos, obtemos a base para a nossa tabela de factos. De seguida, explicamos a transformação em modelo dimensional.

### **7.5.3 Modelo Multidimensional**

Conforme mencionado anteriormente, pretendemos implementar um modelo multidimensional em estrela.

A partir dos registos fornecidos pela GM, extraímos as dimensões Cliente, Centro, Código Postal, *Status* (Estado) e Idade. Todas estas dimensões são conjuntos de dados finitos e determinados. Para cada um destes conjuntos foi criada uma tabela com um campo de identificação unívoco a cada entrada da tabela. Este campo é designado por chave primária da tabela. Não foi possível obter as dimensões Carteiro e Zona, uma vez que estes dados não foram facultados pela GM. A criação da informação para a tabela da dimensão Data foi processada separadamente e é descrita em 7.5.4.

A Figura 7.2 ilustra a tabela da dimensão Cliente. A entrada 1 desta tabela irá definir todos os atributos do cliente. Neste caso definimos apenas o nome do cliente.

	A	B
1	id_cliente	nome_cliente
2	1	A
3	2	B
4	3	C
5	4	D
6	5	E
7	6	F

Figura 7.2 Excerto da Tabela da Dimensão Cliente

A ligação à tabela de factos é feita através da chave primária. No campo cliente, cada entrada em que o cliente seja o cliente “A”, é registada a chave primária que corresponde a esse cliente na tabela da dimensão cliente. Neste caso, 1. Esta relação em que uma entrada numa tabela é invocada diversas vezes noutra tabela é designada de uma relação “um para muitos”. Este registo de uma chave primária noutra tabela é designado por chave estrangeira.

Este processo foi replicado nas restantes dimensões. No final obtemos o esquema em estrela apresentado na Figura 7.3.

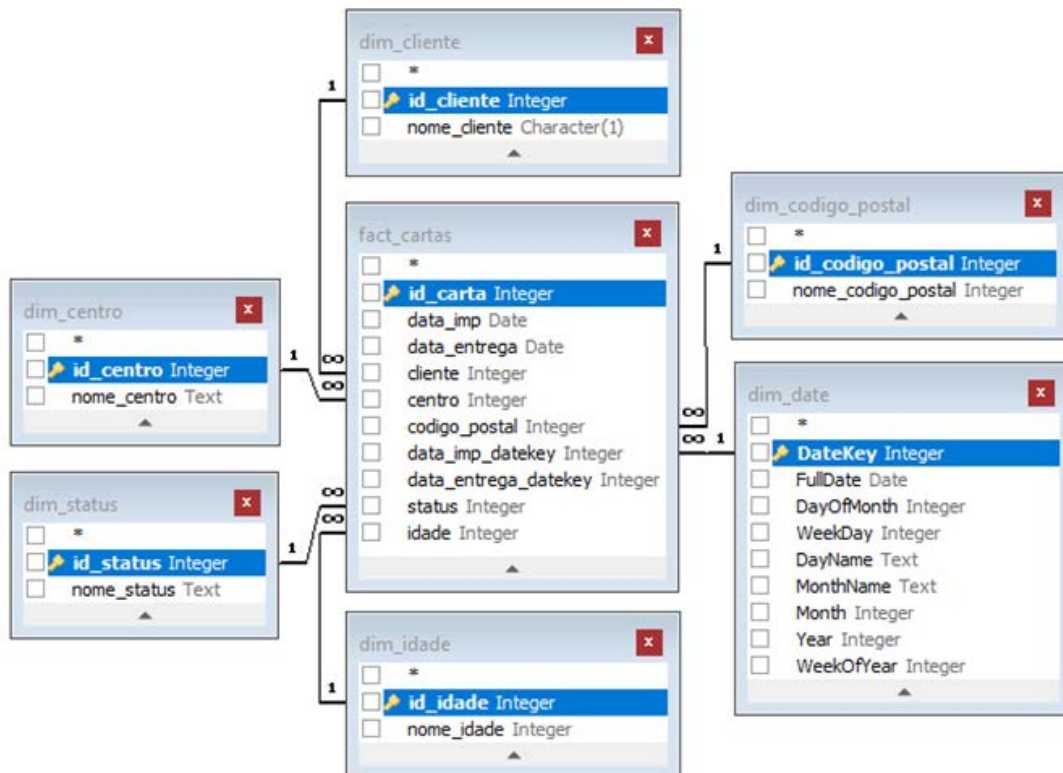


Figura 7.3 Modelo Multidimensional em estrela

### 7.5.4 Dimensão Data Calendário

Na Figura 7.3 anterior podemos observar a tabela da dimensão data de calendário. Esta dimensão assume um papel importante numa plataforma de *Business Intelligence* pois permite análises temporais. Segundo Kimball (2013), está relacionada virtualmente com todas as tabelas de factos por forma a permitir uma navegação através de datas familiares, número da semana, meses, períodos fiscais, e dias especiais do calendário. Kimball sugere que a chave primária seja algo com mais significância como um número inteiro com o formato AAAAMMDD (AnoMêsDia) em vez de um número atribuído de forma sequencial.

Por forma a implementar esta metodologia foram criados dois campos na tabela de factos: `data_imp_datekey` e `data_entrega_datekey`. Estes campos foram preenchidos com um número inteiro no formato AAAAMMDD, convertido a partir da data contida nos campos `data_imp` e `data_entrega` respetivamente.

Seguindo a metodologia de Kimball, foi criada manualmente a tabela para a dimensão Data, com entradas de 1 de abril 2018 a 31 de maio 2018, gerando a informação referente a cada entrada para campos como dia do mês, dia da semana, mês, nome do mês, ano, semana do ano. A Figura 7.4 ilustra um excerto da tabela da dimensão data.

	A	B	C	D	E	F	G	H	I
1	DateKey	FullDate	DayOfMonth	WeekDay	DayName	MonthName	Month	Year	WeekOfYear
2	20180401	01/04/2018	1	1	Domingo	Abril	4	2018	14
3	20180402	02/04/2018	2	2	Segunda	Abril	4	2018	14
4	20180403	03/04/2018	3	3	Terca	Abril	4	2018	14
5	20180404	04/04/2018	4	4	Quarta	Abril	4	2018	14
6	20180405	05/04/2018	5	5	Quinta	Abril	4	2018	14
7	20180406	06/04/2018	6	6	Sexta	Abril	4	2018	14
8	20180407	07/04/2018	7	7	Sabado	Abril	4	2018	14

**Figura 7.4** Excerto da Tabela da Dimensão Data

## 7.6 Instalação da plataforma Pentaho

Com vista ao desenvolvimento prático do projeto, foram identificados e instalados todos os componentes necessários.

Assim, ao nível da plataforma Pentaho foram instalados o servidor, componente principal da plataforma e adicionalmente o *Pentaho Report Designer*, que permite construir relatórios e publicá-los diretamente no Pentaho, e o *Pentaho Schema Workbench* que permite criar cubos OLAP e publicá-los no Pentaho para que possam posteriormente ser utilizados em análises multidimensionais.



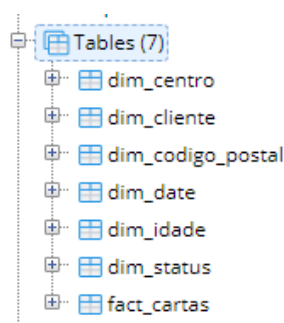
Ao nível do Sistema de Gestão de Base de Dados foi instalado e utilizado o PostgreSQL e o pgAdmin que permite gerir todo o sistema PostgreSQL.

Adicionalmente, foi instalado o *plugin* Saiku no Pentaho. O Saiku (2018) é uma solução que permite aos utilizadores efetuar análises multidimensionais sobre os dados de uma forma fácil e simples. O Saiku liga-se a um espectro de servidores OLAP, incluindo o Mondrian presente no Pentaho. Permite aos seus utilizadores escolher as medidas e dimensões que necessitam analisar e dá acesso interativo a um vasto conjunto de visualizações para que escolham a mais adequada.

Todo o processo de preparação da plataforma pode ser consultado no Anexo 3

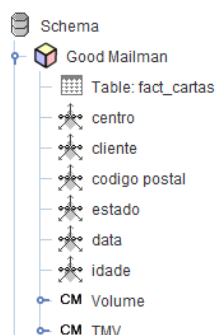
## **7.7 Implementação do Modelo Multidimensional e cubo OLAP**

Uma vez definido o modelo multidimensional, este foi implementado na base de dados com a criação de todas as tabelas de dimensões e tabela de factos (ver Figura 7.5). O código SQL referente à criação de cada tabela pode ser consultado no Anexo 4.



**Figura 7.5** Tabelas criadas em PostgreSQL

Uma vez criadas as tabelas, procedeu-se ao carregamento dos dados em cada uma delas. Terminado este processo, prosseguiu-se para a criação do cubo OLAP com recurso ao *Pentaho Schema Workbench* (PSW). É associada a tabela de factos e criadas as dimensões. É também neste ponto que são criadas as medidas Volume e Tempo Médio de Vida (TMV). A Figura 7.6 ilustra o cubo final. Concluída a configuração do cubo, o PSW permite efetuar o *upload* direto para o Pentaho, ficando de imediato disponível dentro da plataforma.



**Figura 7.6** Cubo contruído com o PSW

Adicionalmente, foi associada uma propriedade de formatação condicional à medida TMV. Esta propriedade é interpretada pelo Saiku, e como veremos adiante, irá permitir introduzir uma formatação condicional nas células de análises em tabelas com valores de TMV. Se a TMV for inferior a 6 dias, a célula toma a cor verde. Se for igual a 6 dias ou superior toma a cor vermelha.

## 7.8 Criação de Análises

Uma vez preparada toda a estrutura de *software*, modelo dimensional implementado e dados carregados para a base de dados, é possível iniciar a criação das análises propostas.

### 7.8.1 Coordenador de Centro

Para a implementação da primeira análise proposta foi escolhido o formato de relatório e no seu desenvolvimento utilizámos o *Pentaho Report Designer* (PRD). O PRD liga-se à base de dados e recolhe os dados necessários à elaboração do relatório através de *queries* em SQL. As *queries* desenvolvidas para a realização deste relatório podem ser consultadas no Anexo 5.

Aproveitando a possibilidade de passagem de parâmetros ao relatório, foi implementada a passagem de dois parâmetros: Mês e Centro. Isto permitirá ao coordenador seleccionar o seu centro e o mês que pretende analisar: um mês anterior ou o atual para analisar o desempenho corrente. A cada parâmetro está associada uma *query* que devolve o conjunto de valores de entrada que o utilizador pode seleccionar. Estas *queries* podem ser consultadas no Anexo 5.

Na Figura 7.7 podemos ver o *layout* do relatório elaborado no PRD com todo o *design* e disposição de campos de dados.


Report Header				
	Análise Mensal / Cliente		MonthName	
	Página: Page			
Group Header	nome_centro			
Group Header	Status: nome_status			
	Cliente	Tempo Médio de Vida	Tempo Max de Vida	Total Cartas
Details	nome cliente	Tempo Médio Vida	Tempo Max Vida	Total Cartas

Figura 7.7 Layout do relatório desenvolvido no PRD

Dentro do relatório, a informação é agrupada. Como podemos ver na Figura 7.7, o primeiro agrupamento é por Centro. Dentro deste grupo (ou relativamente ao Centro), temos o agrupamento seguinte por Estado, tal como definido na proposta. Significa isto que a informação configurada nos detalhes irá ser disponibilizada em dois grupos: “OK” e “Não OK” com definido durante o tratamento de dados.

Podemos ver na Figura 7.8 o resultado obtido quando geramos o relatório seleccionando o centro de Coimbra e o mês de abril. Observamos o agrupamento por estado e nos detalhes a informação é fornecida por cliente conforme proposto.


			
Análise Mensal / Cliente	Abril		
Página: 1 / 1			
Coimbra			
Status: Not OK			
Cliente	Tempo Médio de Vida	Tempo Max de Vida	Total Cartas
M	21	34	10
W	32	45	150
Status: OK			
Cliente	Tempo Médio de Vida	Tempo Max de Vida	Total Cartas
A	8	11	8
B	7	17	8.634
C	6	6	1
D	7	11	626
-	-	-	-

Figura 7.8 Exemplo de relatório por cliente

Constatamos por análise deste relatório que a 17 de maio os clientes M e W teriam ainda cartas sem um estado final, o que constitui um alerta ao coordenador do centro para um problema a ser resolvido.

Além do tempo médio de vida e do volume (total de cartas), foi adicionada uma terceira medida a este relatório que informa sobre a idade máxima que uma carta apresenta no conjunto de dados em análise.

Uma vez terminados e testados, os relatórios podem ser publicados diretamente para o Pentaho. Já na plataforma, podem ser consultados pelos utilizadores e como ilustra a Figura 7.9, antes de gerar o relatório, devem escolher o centro e o mês que pretendem consultar bem como o formato de visualização.

Análise Mensal Centro Código Postal

Row Limit: Maximum

Selecione Centro: Coimbra

Selecione o mês: Abril

Output Type: PDF

View Report  Auto-Submit

**Figura 7.9** Interface de passagem de parâmetros ao relatório

Selecionado o mês e o centro para o relatório que devolve a análise por códigos postais, o resultado foi o apresentado na Figura 7.10.

**GOOD MAILMAN**

Análise Mensal / Código Postal Abril Página: 1 / 1

**Coimbra**

Status: Not OK

Código Postal	Tempo Médio de Vida	Tempo Max de Vida	Total Cartas
3000	32	45	94
3020	30	45	13
3025	25	34	2
3030	31	45	37
3040	20	20	1
3150	30	45	14

Status: OK

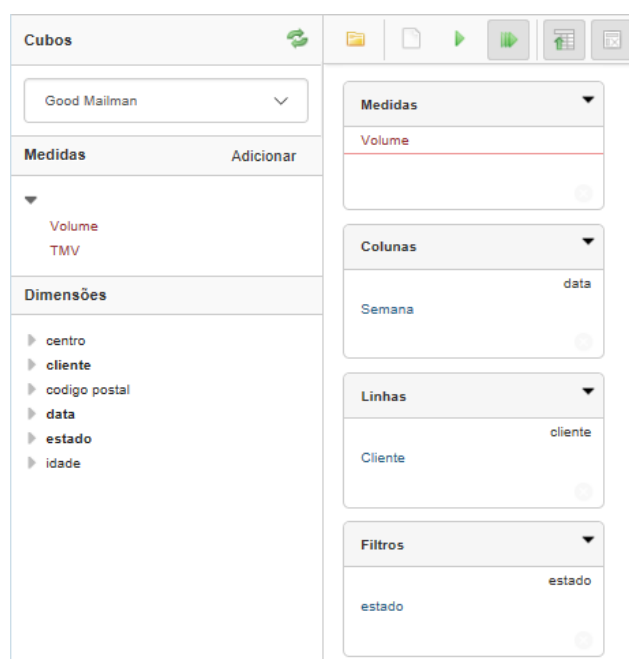
Código Postal	Tempo Médio de Vida	Tempo Max de Vida	Total Cartas
3000	6	29	7.617
3020	6	28	5.146
3025	5	28	4.014
3030	6	28	11.163
3040	5	28	4.259
3045	6	28	2.460
3150	6	23	3.416

**Figura 7.10** Exemplo de relatório por código postal

Apesar de proposta uma análise por Carteiro, tal não foi possível, uma vez que esta informação não foi fornecida pela Good Mailman.

### 7.8.2 Coordenação Geral

Para a implementação da segunda proposta optámos pela utilização do *plugin* Saiku. Como referido antes, este *plugin* permite análises multidimensionais através da ligação a um cubo OLAP. Assim, podemos observar na Figura 7.11 , a interface do Saiku com o cubo desenvolvido anteriormente selecionado, e as medidas e dimensões disponíveis.



**Figura 7.11** Interface do Saiku com ligação ao cubo Good Mailman ativa

Arrastando as dimensões que queremos para as colunas e/ou linhas da análise e a medida pretendida, criamos rapidamente uma análise. Podemos também utilizar dimensões para criar filtros. No exemplo da Figura 7.11 acima, vamos analisar a evolução semanal do volume de cartas por cliente. No filtro foi definido que pretendíamos apenas o conjunto de cartas com estado “OK”. O resultado obtido foi o apresentado na Figura 7.12.

Tal como definido no cubo, temos aqui representado através da formatação condicional o indicador da qualidade de serviço referente ao tempo médio de vida. Podemos facilmente verificar que os clientes W e X, apesar de um volume considerável, mantiveram ao longo de todas as semanas em análise um tempo médio de vida inferior a 6 dias (verde).

data - Semana	14	15	16	17	18	19	20
Cliente	Volume	Volume	Volume	Volume	Volume	Volume	Volume
A	2	3	5	6	1	2	1
B	2391	4743	4712	4880	4745	689	-
C	-	-	1	-	1	-	-
D	54	191	190	211	190	22	1
E	-	3	1	-	1	1	-
F	-	1	3	1	1	-	-
G	8	-	2	-	8	-	-
H	282	602	468	708	649	138	1
J	-	1247	1324	1041	-	781	-
K	54	-	-	-	59	-	-
L	9	-	-	-	12	-	-
M	47	177	113	51	10	1	-
N	4	-	1	3	-	-	-
O	1634	8009	4682	544	1322	1169	-
R	11587	8813	9304	4975	4085	2846	-
V	-	2	-	-	-	-	-
W	10357	4463	1482	1250	699	8635	665
X	3497	4508	3629	3473	3420	3378	86

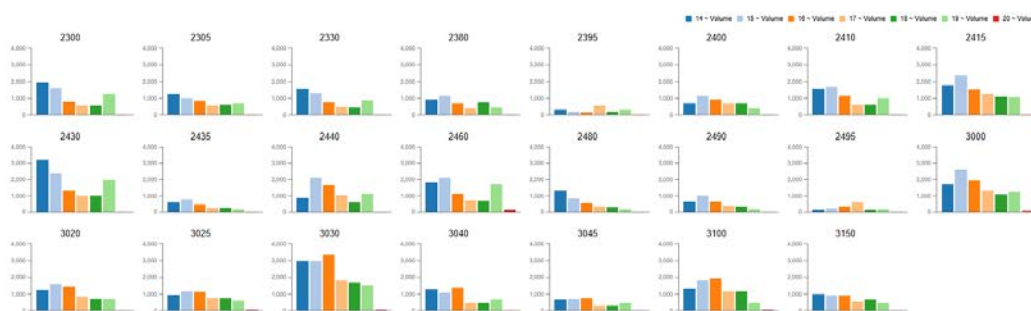
**Figura 7.12** Análise da evolução semanal por cliente obtida com recurso ao Saiku

A mesma análise, mas por código postal, como ilustra a Figura 7.13 dá-nos uma visão diferente e permite perceber que na semana 17, de uma forma geral, o tempo médio de vida ultrapassou os 5 dias, mas que o desempenho foi recuperado entre a semana 18 e 19. Na semana 17 ocorreu o feriado do 25 de abril que levou a um dia de paragem que pode ter tido algum impacto neste aspeto.

data - Semana	14	15	16	17	18	19	20
Código Postal	Volume	Volume	Volume	Volume	Volume	Volume	Volume
2300	1943	1604	811	602	603	1253	3
2305	1273	1032	854	600	622	696	3
2330	1576	1321	738	503	436	874	3
2380	916	1128	728	394	776	473	-
2395	338	212	152	601	179	312	-
2400	730	1158	930	711	694	421	-
2410	1568	1691	1143	641	611	1003	14
2415	1796	2389	1533	1258	1118	1038	40
2430	3193	2371	1320	1001	1008	1955	48
2435	609	773	488	241	246	153	-
2440	887	2097	1662	1057	585	1140	-
2460	1817	2115	1115	754	675	1702	117
2480	1294	834	563	321	296	157	-
2490	650	1006	657	363	354	176	1
2495	128	187	356	618	125	149	8
3000	1701	2649	1937	1289	1104	1223	86
3020	1299	1601	1416	845	699	697	30
3025	974	1153	1130	747	755	598	70
3030	2991	2957	3365	1808	1704	1531	75
3040	1294	1068	1374	506	487	651	39
3045	677	699	758	316	303	490	29
3100	1324	1803	1963	1156	1158	472	70
3150	980	914	925	590	646	477	19

**Figura 7.13** Análise da evolução semanal por código postal obtida com o Saiku

O Saiku permite alterar rapidamente o tipo de visualização usada e a análise anterior pode facilmente ser, por exemplo, analisada em gráfico de barras de forma individual para cada código postal como ilustra a Figura 7.14.



**Figura 7.14** Evolução semanal em Gráfico de Barras por Código Postal com o Saiku

Após obtida a análise pretendida, o utilizador pode guardá-la na sua área de trabalho por forma a permitir posteriormente uma consulta mais rápida.

### **7.8.3 Dashboard**

Na implementação do *dashboard* foi utilizado o *Community Dashboard Editor (CDE)*. Este *plugin* é livre e vem pré-instalado na plataforma Pentaho. O desenho do *dashboard* com o CDE rege-se segundo uma matriz de linhas e colunas onde definimos onde queremos colocar painéis com informação.

Seguindo o objetivo, foram criados dois painéis para o topo do *dashboard* com a evolução semanal do TMV das cartas com estado OK e outro para a evolução do volume total de cartas. Abaixo do painel com a evolução semanal do TMV, seguindo esta orientação, foram colocados dois painéis com a evolução do TMV por Código Postal e por Cliente, adicionando algum detalhe à informação que o painel superior confere.

Para o lado direito inferior, mantendo a relação com o painel superior, foram criados dois painéis com a evolução de volume por centros. Um com a informação relativa ao Volume com estados finais (OK) e outro com informação relativa a estados não finais (Não OK). O *dashboard* final pode ser observado na Figura 7.15.

A informação que cada painel disponibiliza é obtida individualmente por meio de *queries MDX* sobre o cubo OLAP desenvolvido anteriormente. Estas *queries* podem ser consultadas no Anexo 6.



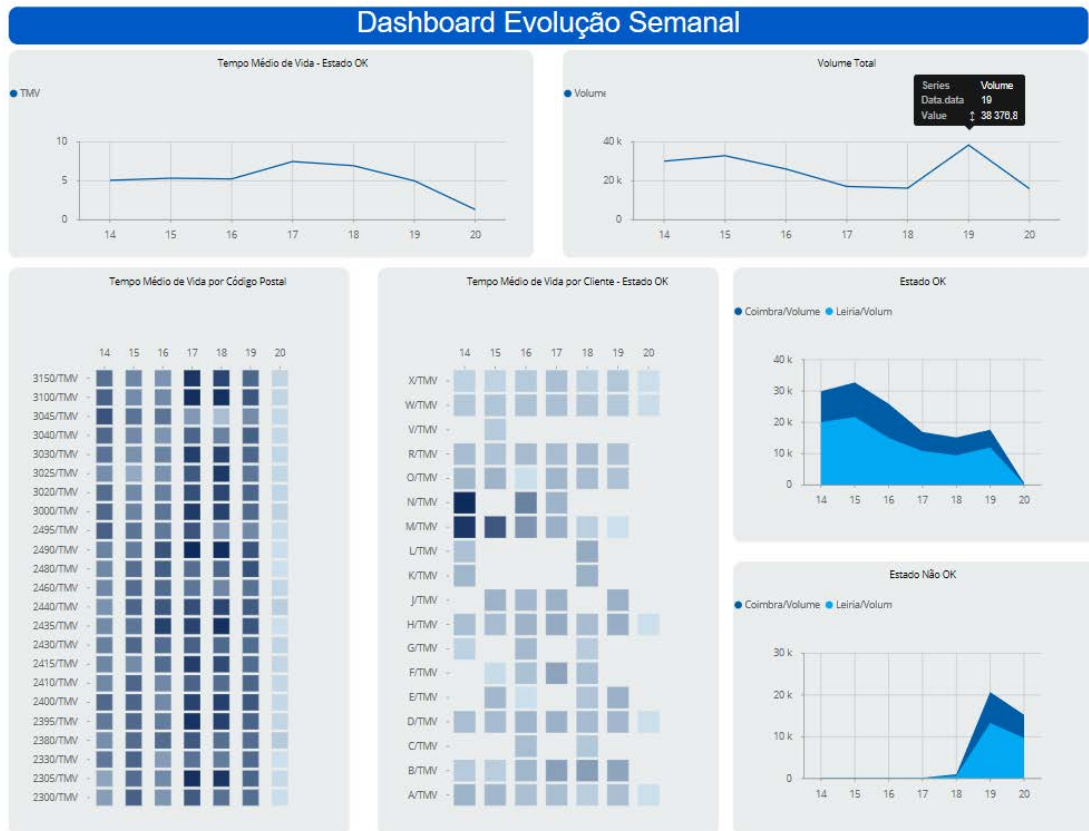


Figura 7.15 Dashboard desenvolvido com o CDE

## 7.9 Conclusões

O anonimato solicitado pela empresa e o conjunto de dados relativamente reduzido, limitou quer a fase de contextualização dos processos operacionais, quer o leque de análises implementadas, quer a informação possível de obter. Porém, tal não limitou o alcançar dos objetivos propostos.

Na obtenção de análises multidimensionais, é de salientar a utilização do *plugin* Saiku que surpreendeu pela sua facilidade de uso e versatilidade em análises e por transmitir poder de análise ao utilizador. É um *plugin* a ter em conta na plataforma Pentaho.

Conforme pretendíamos demonstrar, o processo de implementação de uma plataforma *open source* de BI como a Pentaho, está perfeitamente ao alcance das organizações. A Pentaho em particular, na sua versão *open source*, ainda que sem as mais recentes funcionalidades apenas presentes na versão comercial, prova ser uma plataforma robusta, fiável, e capaz acrescentar valor às organizações que a escolherem como solução de *Business Intelligence*. A par de uma boa documentação de suporte, a comunidade de utilizadores da versão *open source* é considerável e uma mais valia quando utilizadores menos experientes necessitam de obter ajuda. No entanto, é de registar que algum



conhecimento prévio nesta área facilitou a implementação prática da plataforma. Significa isto que as organizações devem ter em conta a necessidade de, pelo menos, algum investimento em suporte técnico, seja ele interno ou externo.

## **8 CONCLUSÕES E TRABALHO FUTURO**

O presente trabalho de projeto teve como motivação principal reforçar a demonstração das soluções *open source* de *Business Intelligence* como uma alternativa viável às soluções comerciais e do *Business Intelligence* como um tema de relevância crescente na atualidade.

Os sistemas de *Business Intelligence* reúnem dados de diversas fontes, processando-os e analisando, gerando informação e conhecimento. Esta informação e conhecimento têm o intuito de auxiliar tomadas de decisão e, desta forma, permitir tomar “a melhor” decisão. Atualmente, tomar “a melhor decisão” tem uma importância elevada dentro das organizações e o *Business Intelligence* como meio facilitador desta tarefa, assume uma crescente relevância.

Atualmente, as vantagens do BI como por exemplo a otimização de recursos, de processos e melhoria da eficácia, servem todo o tipo de organizações, tenham elas fins lucrativos ou não. Desde que estas organizações disponham de dados operacionais capazes de “alimentar” um sistema de BI, estão potencialmente aptas a tirar partido de todas essas vantagens.

O custo com licenças em soluções comerciais não está ao alcance de todas as organizações. Este é um dos fatores para a não adoção de uma solução de BI dentro de uma empresa e também o primeiro a apontar para as soluções *open source* como alternativa.

A escolha do tema *Business Intelligence* para este trabalho e em particular, soluções *open source*, constitui por si só um contributo para a crescente relevância do BI e das soluções *open source* na comunidade.

Iniciou-se este trabalho com uma revisão da literatura sobre *Business Intelligence* e sistemas de BI, não esquecendo também de abordar a importância do BI na tomada de decisão. São vários os autores que referem o BI como uma peça chave na tomada de decisão. E tomadas de decisão a vários níveis: estratégico, tático e operacional (Vercellis, 2009).

Seguiu-se a identificação das principais plataformas de BI atualmente disponíveis, em desenvolvimento, e neste processo destacaram-se a Jaspersoft, Knowage e Pentaho.

Numa primeira avaliação, estas plataformas foram comparadas entre si pela validação de critérios. Estes critérios basearam-se nos critérios do Quadrante Mágico de Gartner (Sallam et al., 2017). Concluiu-se nesta avaliação que a Knowage validou mais critérios que a Jaspersoft e a Pentaho. Tanto a Jaspersoft como a Pentaho obtiveram o mesmo resultado, ocupando a segunda posição.

Assim, e seguindo com o objetivo de selecionar a plataforma que atualmente apresenta mais potencial para implementação, realizámos uma nova avaliação, desta vez com recurso a uma metodologia de avaliação de *software open source*. Foram submetidas a esta segunda avaliação a plataforma Knowage (tendo sido a melhor na avaliação anterior) e a Pentaho. Num conjunto de artigos de comparação de plataformas *open source* de BI, a Pentaho obtém reiteradamente melhores resultados que a Jaspersoft, motivo pelo qual foi a segunda plataforma escolhida.

Nesta avaliação foi utilizada a metodologia OSSpal. Entre várias metodologias com mais de uma década, esta surge destacada com uma revisão e atualização recente. Nesta avaliação, a Pentaho obteve melhor resultado que a Knowage mas ainda assim próximas. Tanto a Knowage com a Pentaho revelam-se boas opções atualmente, sendo que a vantagem da Pentaho está sobretudo numa maior comunidade de suporte e mais e melhor documentação, aspetos importantes na adoção de *software open source*. Prova dada da sua qualidade, está também no nível de implementação a nível mundial, que é significativo. O facto de ter uma boa documentação de suporte veio, efetivamente, a revelar-se importante na demonstração prática.

Uma forma complementar de divulgar as soluções *open source* de BI como alternativas viáveis, é demonstrar, de forma prática, a sua implementação e utilização. Neste sentido, o Caso de Estudo Good Mailman pretendeu sobretudo demonstrar que a utilização de uma plataforma *open source* de BI não é um processo complicado e que está ao alcance das organizações. Para o efeito, foi desenvolvida uma prova de conceito, recorrendo a dados de uma empresa portuguesa. Contextualizou-se a empresa, o negócio e a parte operacional a ser analisada. Definiram-se objetivos em termos da informação que se pretendia obter dos dados fornecidos.

Tratados os dados, e aplicados alguns conceitos e processos de BI, foi possível produzir relatórios, análises multidimensionais e um *dashboard* que deram resposta aos objetivos traçados.

Nas análises multidimensionais foi utilizado o *plugin* Saiku na sua versão open source que surpreendeu pela sua facilidade de uso, versatilidade em análises e por transmitir poder de análise ao utilizador.

Com o Saiku, o utilizador cria as suas análises multidimensionais de forma simples e rápida, podendo guardar as análises na sua área de trabalho individual para consulta futura. É assim um *plugin* de grande valor dentro da plataforma Pentaho e que potencia a vertente *Self-Service* BI em grande crescimento atualmente.

É neste ponto, de mencionar, que algum conhecimento prévio na área de Bases de Dados e linguagem SQL, bem como alguma experiência com um *software* semelhante ao PRD, facilitou esta demonstração prática. Significa isto que uma empresa que pretenda implementar uma plataforma de BI terá sempre de ter em atenção a importância de ter algum suporte técnico, seja ele interno ou externo, sendo que esta condição pode implicar alguns custos. Ainda assim, os custos associados a esta condição não são, porém, comparáveis aos valores que as soluções comerciais apresentam apenas para licenciamento.

Uma dificuldade sentida na parte prática foi a limitação imposta pelo anonimato e o conjunto de dados fornecido ser reduzido. Não inviabilizando o processo, limitou, por um lado, a análise dos processos operacionais da empresa e, por outro, o conjunto de análises possíveis de efetuar.

Concluimos que a crescente importância do *Business Intelligence* na sociedade como um todo é atualmente indiscutível e deve ser um ponto na agenda de todas as organizações. As soluções *open source* são sem dúvida uma alternativa viável. Como ficou demonstrado, a Pentaho deu uma resposta integral aos problemas propostos, de uma forma simples, sem requerer um conhecimento profundo das tecnologias envolvidas. É uma plataforma com um espectro de utilização mundial superior à Knowage, o que lhe confere um estatuto de *software* com provas dadas e aceitação por parte do mercado. Tem uma vasta comunidade de utilizadores que, por um lado, potencia o aparecimento e evolução de *plugins*, e por outro, constitui um importante grupo de suporte a todos os utilizadores menos experientes. A Pentaho, é assim, a que atualmente recomendamos às organizações interessadas em adotar este caminho do *open source*.

No decorrer deste trabalho, foram elaborados dois artigos, submetidos e aceites em conferências:

- “Avaliação Comparativa de Plataformas de *Business Intelligence Open Source* para PME”, CISTI – 13ª Conferência Ibérica de Sistemas e Tecnologias de Informação, Cáceres, Espanha, Junho 2018 (Anexo 1)
- “*Open Source Business Intelligence Platforms’ Assessment using OSSpal Methodology*”, ICE-B – *International Conference on e-Business*, Porto, Portugal, Julho 2018 (Anexo 2).

A publicação destes artigos contribui para a divulgação das soluções *open source* de BI bem como auxilia as organizações que pretendam implementar uma solução de BI e estejam no processo de escolha.

Por fim, a nível pessoal, destaco a experiência enriquecedora que esta análise de soluções *open source* no âmbito do *Business Intelligence* representou e que em muito contribuiu para a tomada de consciência neste tema.

Como trabalho futuro, é importante seguir o acompanhamento da evolução destas plataformas *open source* de *Business Intelligence* uma vez que neste momento todas têm uma versão comercial e as funcionalidades disponíveis nas versões *open source* podem ser alteradas. Também importante será um exemplo prático de implementação em ambiente empresarial, sem restrições de acesso aos dados, permitindo uma melhor demonstração das capacidades destas plataformas.

## **REFERÊNCIAS BIBLIOGRÁFICAS**

- Ahishakiye, E., Omulo, E. O., Taremwa, D., & Wario, R. (2017). Comparative Analysis of Open source Business Intelligence tools for Crime Data Analytics. *International Journal of Latest Research in Engineering and Technology (IJLRET)*, 03(04), 60–65.
- Al-Debei, M. M. (2011). Data Warehouse as a Backbone for Business Intelligence: Issues and Challenges. *European Journal of Economics, Finance & Administrative Sciences*, 33(33), 153–166.
- Ariyachandra, T., & Watson, H. (2010). Key organizational factors in data warehouse architecture selection. *Decision Support Systems*, 49(2), 200–212. <https://doi.org/10.1016/j.dss.2010.02.006>
- Barros, R. M. P. da C. (2013). Dashboarding - Projeto e Implementação de Painéis Analíticos, 1–91.
- Bernardino, J., & Tereso, M. (2013). Business Intelligence Tools. In *Computational Intelligence and Decision Making* (pp. 267–276). Springer, Dordrecht. [https://doi.org/10.1007/978-1-4020-5614-7\\_301](https://doi.org/10.1007/978-1-4020-5614-7_301)
- Berry, M. J. a., & Linoff, G. S. (2004). *Data mining techniques: for marketing, sales, and customer relationship management* (2<sup>a</sup> Ed). Wiley Publishing, Inc. Retrieved from <http://portal.acm.org/citation.cfm?id=983642>
- Bertolini, A., Chiappin, M., Mayolo, V., D'Arrigo, F., Barcellos, P., & Dias, D. (2015). Soluções Business Intelligence Open Source no Suporte à Estratégia Organizacional. *Revista Inteligência Competitiva*, 40–59. <https://doi.org/http://dx.doi.org/10.24883%2Fric.v5i2.116>
- Brandão, A., Pereira, E., Esteves, M., Portela, F., Santos, M. F., Abelha, A., & Machado, J. (2016). A benchmarking analysis of open-source business intelligence tools in healthcare environments. *Information — Open Access Journal of Information Science*, 7(4), 1–16. <https://doi.org/10.3390/info7040057>
- Chaudhary, S. (2004). Management Factors for Strategic BI Success. *Business Intelligence in Digital Economy. Opportunities, Limitations and Risks*.
- Chaudhuri, S., & Dayal, U. (1997). An overview of data warehousing and OLAP technology. *ACM SIGMOD Record*, 26(1), 65–74.

- <https://doi.org/10.1145/248603.248616>
- Chaudhuri, S., Dayal, U., & Narasayya, V. (2011). An overview of business intelligence technology. *Communications of the ACM*, 54(8), 88. <https://doi.org/10.1145/1978542.1978562>
- Costa, S. (2012). *Sistema de Business Intelligence como suporte à Gestão Estratégica*. Universidade do Minho. Retrieved from <http://hdl.handle.net/1822/25810>
- Davenport, T. H. (2006). Competing on Analytics. *Harvard Business Review*, 84(Jan.), 98–107.
- Deprez, J. C., & Alexandre, S. (2008). Comparing Assessment Methodologies for Free/Open Source Software: OpenBRR and QSOS. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5089 LNCS, 189–203. [https://doi.org/10.1007/978-3-540-69566-0\\_17](https://doi.org/10.1007/978-3-540-69566-0_17)
- Eckerson, W. (2005). *Performance Dashboards: Measuring, Monitoring, and Managing Your Business*. John Wiley & Sons, Inc.
- Engineering, G. (2018). Engineering Group. Retrieved January 29, 2018, from <http://www.eng.it/>
- Ferreira, M., Ferros, L. M., & Fernandes, V. (2012). Avaliação e seleção de software open-source para Gestão Integrada de Bibliotecas. *Actas Do Congresso Nacional de Bibliotecários, Arquivistas e Documentalistas*, 6. Retrieved from <http://repositorium.sdum.uminho.pt/handle/1822/20491>
- Ferreira, M., Silva, R., Vieira, V., Guimarães, C., & Carvalho, J. (2010). Um estudo de caso com análise comparativa entre ferramentas de BI livre e proprietária. Retrieved from [https://www.researchgate.net/profile/Vaninha\\_Vieira/publication/266492563\\_Um\\_estudo\\_de\\_caso\\_com\\_analise\\_comparativa\\_entre\\_ferramentas\\_de\\_BI\\_livre\\_e\\_proprietaria/links/55397b550cf247b858812e5a.pdf](https://www.researchgate.net/profile/Vaninha_Vieira/publication/266492563_Um_estudo_de_caso_com_analise_comparativa_entre_ferramentas_de_BI_livre_e_proprietaria/links/55397b550cf247b858812e5a.pdf)
- Ferreira, T., Pedrosa, I., & Bernardino, J. (2017). Evaluating Open Source Business Intelligence Tools using OSSpal Methodology. In *Proceedings of the 9th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management* (pp. 283–288).

- <https://doi.org/10.5220/0006516402830288>
- Few, S. (2004). Dashboard Confusion. Retrieved June 7, 2018, from [http://www.perceptualedge.com/articles/ie/dashboard\\_confusion.pdf](http://www.perceptualedge.com/articles/ie/dashboard_confusion.pdf)
- Few, S. (2006). *Information Dashboard Design: The Effective Visual Communication of Data*. O'Reilly.
- Gaardboe, R., & Svarre, T. (2017). Critical Factors for Business Intelligence Success. In *25th European Conference on Information Systems (ECIS)* (Vol. 2017). Guimarães. Retrieved from [https://aisel.aisnet.org/ecis2017\\_rp/31](https://aisel.aisnet.org/ecis2017_rp/31)
- Golfarelli, M. (2009). Open source bi platforms: A functional and architectural comparison. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5691 LNCS, 287–297. [https://doi.org/10.1007/978-3-642-03730-6\\_23](https://doi.org/10.1007/978-3-642-03730-6_23)
- Golfarelli, M., Maio, D., & Rizzi, S. (1998). Conceptual design of data warehouses from E/R schemes. In *Proceedings of the Thirty-First Hawaii International Conference on System Sciences* (Vol. 7, pp. 334–343). <https://doi.org/10.1109/HICSS.1998.649228>
- Habul, A., & Pilav-Velic, A. (2010). Business Intelligence and Customer Relationship Management. In *Proceedings of the ITI 2010 32nd Int. Conf. on Information Technology Interfaces* (pp. 169–174).
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques*. San Francisco, CA, itd: Morgan Kaufmann (3<sup>a</sup> Ed). <https://doi.org/10.1016/B978-0-12-381479-1.00001-0>
- Inmon, W. H. (2005). *Building the Data Warehouse*. (Wiley, Ed.), *The Encyclopedia of Data Warehousing and Mining* (4<sup>a</sup> Ed., Vol. 13).
- Jaspersoft. (2018a). Arquitetura Jaspersoft. Retrieved February 16, 2018, from <https://www.jaspersoft.com/es/jaspersoft-inicio-rapido-guia>
- Jaspersoft. (2018b). Interface Jaspersoft Reports. Retrieved from <https://mobiledemo.jaspersoft.com/jasperserver-pro/login.html>
- Jaspersoft. (2018c). Jaspersoft Web Page. Retrieved January 24, 2018, from <https://www.jaspersoft.com/>



- Kimball, R., & Ross, M. (2002). *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*. (R. Elliott & E. Herman, Eds.) (2nd Ed.). Wiley Computer Publishing.
- Kimball, R., & Ross, M. (2013). *The Data Warehouse Tollkit: The Definitive Guide to Dimentional Modeling* (3<sup>a</sup> Ed.). John Wiley & Sons, Inc.
- Knowage. (2018a). Knowage Web Page. Retrieved January 24, 2018, from <https://www.knowage-suite.com/site/home/>
- Knowage. (2018b). Relatório Knowage. Retrieved from <https://www.knowage-suite.com/site/product/enterprise-reporting/>
- Lapa, J., Bernardino, J., & Figueiredo, A. (2014). A comparative analysis of open source business intelligence platforms. In *Proceedings of the ISDOC '14* (pp. 86–92). <https://doi.org/10.1145/2618168.2618182>
- Larson, B. (2009). *Delivering Business Intelligence with Microsoft SQL Server 2008*. The McGraw-Hill Companies, Inc.
- Madewell, C. (2014). Advantages & Disadvantages of Implementing Business Intelligence. Retrieved from <https://www.linkedin.com/pulse/advantages-disadvantages-implementing-business-charles-d-madewell/>
- Maimon, O., & Rokach, L. (2010). *Data Mining and Knowledge Discovery Handbook*. Springer (2<sup>a</sup> Ed.). Springer Science Business Media, Inc. <https://doi.org/10.1007/978-0-387-09823-4>
- Malik, S. (2005). *Enterprise Dashboards: Design and Best Pratices fo IT*. John Wiley and Sons Inc (Vol. 1). John Wiley & Sons, Inc.
- Marinheiro, A., & Bernardino, J. (2013). OpenBRR evaluation of an open source BI suite. *Proceedings of the International C\* Conference on Computer Science and Software Engineering - C3S2E '13, 1*, 134. <https://doi.org/10.1145/2494444.2494463>
- Marinheiro, A., & Bernardino, J. (2015). Experimental Evaluation of Open Source Business Intelligence Suites using OpenBRR. *IEEE Latin America Transactions*, 13(3), 810–817. <https://doi.org/10.1109/TLA.2015.7069109>
- Negash, S. (2004). Business intelligence. *Communications of the Association for Information Systems*, Vol.13(January), 177–195.

- Nereu, F. (2017). *Open Source Platforms for Big Data Analytics*.
- Olszak, C. M., & Ziemba, E. (2004). Business intelligence systems as a new generation of decision support systems. In *PISTA 2004, International Conference on Politics and Information Systems: Technologies and Applications*.
- Olszak, C. M., & Ziemba, E. (2007). Approach to building and implementing Business Intelligence systems. *Interdisciplinary Journal of Information, Knowledge, and Management*, 2, 135–148.
- OpenBRR. (2005). OpenBRR White paper. Retrieved April 12, 2018, from [https://web.archive.org/web/20050803022846/http://www.openbrr.org/docs/BRR\\_whitepaper\\_2005RFC1.pdf](https://web.archive.org/web/20050803022846/http://www.openbrr.org/docs/BRR_whitepaper_2005RFC1.pdf)
- Parra, V. M., & Syed, A. (2016). Pentaho and Jaspersoft : A Comparative Study of Business Intelligence Open Source Tools Processing Big Data to Evaluate Performance, (November).
- Pentaho. (2018a). Arquitetura Pentaho. Retrieved February 16, 2018, from <https://forums.pentaho.com/threads/149369-Pentaho-5-0-blends-right-in!>
- Pentaho. (2018b). Pentaho Web Page. Retrieved January 24, 2018, from <http://www.pentaho.com/>
- Pentaho. (2018c). Relatório Pentaho. Retrieved from <http://www.pentaho.com/product/business-analytics>
- Petrinja, E., Sillitti, A., & Succi, G. (2008). Overview on trust in large FLOSS communities. *IFIP International Federation for Information Processing*, 275, 47–56. [https://doi.org/10.1007/978-0-387-09684-1\\_4](https://doi.org/10.1007/978-0-387-09684-1_4)
- Petrinja, E., Sillitti, A., & Succi, G. (2010). Comparing OpenBRR, QSOS, and OMM Assessment Models. In *IFIP International Federation for Information Processing* (pp. 224–238). [https://doi.org/https://doi.org/10.1007/978-3-642-13244-5\\_18](https://doi.org/https://doi.org/10.1007/978-3-642-13244-5_18)
- Piedade, M. (2011). Business Intelligence no suporte ao conceito e à prática de Student Relationship Management em instituições de Ensino Superior, 286. Retrieved from <http://hdl.handle.net/1822/20461>
- Power, D. J. (2007). A Brief History of Decision Support Systems. Retrieved from <http://dssresources.com/history/dsshistory.html>

- Rainardi, V. (2008). *Building a Data Warehouse: With Examples in SQL Server*. Apress.
- Reinschmidt, J., & Francoise, A. (2000). Business Intelligence Certification Guide. *IBM International Technical*, 166. Retrieved from <http://www.redbooks.ibm.com/redbooks/pdfs/sg245747.pdf>
- Ribeiro, P., & Bernardino, J. (2010). Analysis of open source Business Intelligence suites. In *E-ALT2010 - E.Activity and Learning Technologies*.
- Saiku. (2018). Saiku. Retrieved May 3, 2018, from <https://www.meteorite.bi/>
- Sallam, R., Howson, C., Idoine, C. J., Oestreich, T. W., Richardson, J. L., & Tapadinhas, J. (2017). Magic quadrant for business intelligence and analytics platforms. *Gartner*, (February 2017), 1–126. <https://doi.org/10.1017/CBO9781107415324.004>
- Samoladas, I., & Gousios, G. (2008). The SQO-OSS quality model: measurement based open source software evaluation. *Open Source Development, Communities and Quality. OSS 2008. IFIP – The International Federation for Information Processing*, 275, 237–248. [https://doi.org/10.1007/978-0-387-09684-1\\_19](https://doi.org/10.1007/978-0-387-09684-1_19)
- Santos, M. Y., & Ramos, I. (2009). *Business Intelligence - Tecnologias da Informação na Gestão de Conhecimento* (2ª ed.). Lisboa: FCA.
- Sassi, R. J. (2010). Data Webhouse E Business Intelligence Operacional : Revisitando a Tecnologia E Analisando As Tendências Do Armazém De Dados. *Encontro Nacional de Engenharia de Produção*.
- Sezões, C., Oliveira, J., & Baptista, M. (2006). *Business Intelligence*. SPI – Sociedade Portuguesa de Inovação. Retrieved from [http://web.spi.pt/negocio\\_electronico/documentos/manuais\\_PDF/Manual\\_V.pdf](http://web.spi.pt/negocio_electronico/documentos/manuais_PDF/Manual_V.pdf)
- Sidiqui, M. A., & Mukhi, S. (2011). Business Intelligence tools ' comparison MS SQL Server Vs Pentaho Open Source Proprietary BI products. *Journal of Independent Studies and Research – Computing*, 9(1).
- SpagoBI, W. (2014). Arquitetura SpagoBI. Retrieved February 16, 2018, from [http://wiki.spagobi.org/xwiki/bin/view/spagobi\\_server/analytical\\_engines](http://wiki.spagobi.org/xwiki/bin/view/spagobi_server/analytical_engines)
- Tereso, M., & Bernardino, J. (2011). Open source business intelligence tools for SMEs. In *6th Iberian Conference on Information Systems and Technologies (CISTI)* (pp. 1–4). Retrieved from

<http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5974187>

- Thomsen, C., & Pedersen, T. B. (2008). A Survey of Open Source Tools for Business Intelligence. *International Journal of Data Warehousing and Mining*, 5(3), 56–75. <https://doi.org/10.4018/jdwm.2009070103>
- Turban, E., Sharda, R., & Delen, D. (2011). *Decision Support and Business Intelligence Systems*. Prentice Hall (9<sup>a</sup> Ed.).
- Vassiliadis, P., Simitsis, A., & Skiadopoulou, S. (2002). Conceptual modeling for ETL processes. *Proceedings of the 5th ACM International Workshop on Data Warehousing and OLAP - DOLAP '02*, 14–21. <https://doi.org/10.1145/583890.583893>
- Velcu-Laitinen, O., & Yigitbasioglu, O. M. (2012). The Use of Dashboards in Performance Management: Evidence from Sales Managers. *The International Journal of Digital Accounting Research*, 12, 39–58.
- Vercellis, C. (2009). *Business intelligence: data mining and optimization for decision making* (1st ed.). John Wiley and Sons, Ltd.
- Wasserman. (2014). OSSpal Model. Retrieved January 12, 2018, from [http://oss.sv.cmu.edu/presentations/WeekThree2014\\_Wasserman.pdf](http://oss.sv.cmu.edu/presentations/WeekThree2014_Wasserman.pdf)
- Wasserman, A. I., Guo, X., Mcmillian, B., & Qian, K. (2017). OSSpal: Finding and Evaluating Open Source Software. In *Open Source Systems: Towards Robust Practices* (Vol. 496, pp. 193–203). <https://doi.org/10.1007/978-3-319-57735-7>



## **ANEXOS**



## **ANEXO 1**

### **Avaliação Comparativa de Plataformas de *Business Intelligence* *Open Source* para PME – CISTI 2018**



# Avaliação Comparativa de Plataformas de Business Intelligence Open Source para PME

## *Comparative Evaluation of Open Source Business Intelligence Platforms for SME*

Nuno Leite<sup>1</sup>, Isabel Pedrosa<sup>2</sup>

Inst. Sup. de Contabilidade e Administração de Coimbra  
Instituto Politécnico de Coimbra  
Coimbra, Portugal

<sup>1</sup>iscac16152@alumni.iscac.pt, <sup>2</sup>ipedrosa@iscac.pt

Jorge Bernardino

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

**Resumo** — O conhecimento que se consegue extrair dos dados existentes nas organizações é fundamental para aumentar a sua vantagem competitiva nos atuais mercados em constante mutação. Neste contexto, a utilização de plataformas de *Business Intelligence* (BI) tem-se mostrado uma escolha eficaz no suporte à tomada de decisão. As plataformas de BI estão a tornar-se um grande ativo para qualquer Pequena e Média Empresa (PME), uma vez que apresentam múltiplos benefícios: uso eficiente de recursos, aumento da quota de mercado, identificação de oportunidades de negócio e de tendências negativas. As plataformas de BI de código aberto oferecem a maioria das funcionalidades disponíveis em soluções comerciais, sem custo de licenciamento, sendo por isso uma alternativa viável para as PME. No entanto, é fundamental que a PME escolha a plataforma de BI mais completa. Neste artigo, com o objetivo de ajudar nesta escolha, foi efetuada uma avaliação comparativa de três plataformas de BI de código aberto mais utilizadas: Jaspersoft, Knowage (SpagoBI) e Pentaho. Nesta avaliação foram usados os critérios da Gartner, tendo a Knowage obtido os melhores resultados.

**Palavras Chave** – *Business Intelligence; Plataformas Open Source; PME.*

**Abstract** — The knowledge one can extract from existing data in organizations is critical to increasing their competitive advantage in today's changing markets. In this context, the use of Business Intelligence (BI) platforms has been shown to be an effective choice in support of decision-making. BI platforms are becoming a major asset for any Small and Medium-Sized Enterprise (SME), as they have multiple benefits: efficient resource use, increased market share, identification of business opportunities and negative trends. Open source BI platforms provide most of the functionalities available in commercial solutions without licensing costs, and is therefore a viable alternative for SME. However, it is critical that SME choose the complete BI platform. In this paper, a comparative assessment was made of three commonly used Open Source BI platforms: Jaspersoft, Knowage (SpagoBI) and Pentaho. In this evaluation were used the criteria of Gartner, and the Knowage obtained the best results.

**Keywords** – *Business Intelligence; Open Source Platforms; SME.*

### I. INTRODUÇÃO

O termo Business Intelligence terá sido introduzido por Howard Dresner do Grupo Gartner em 1989 [1]. Davenport em [2] define um sistema de BI como um conjunto de processos e de *software* utilizado para recolher, analisar e disseminar dados, com o intuito de uma melhor tomada de decisão.

*Business Intelligence* traduz-se num conjunto de práticas de gestão, implementadas através de *software*, com o objetivo de aumentar a rentabilidade e apoiar as administrações na tomada de decisão e liderança das suas organizações [3]. As empresas que recorrem a plataformas de BI, têm ao dispor ferramentas analíticas que fornecem informação e dados importantes para a sua gestão.

As pequenas e médias empresas (PME) desempenham um papel importante no desenvolvimento socioeconómico e são a maior força na prosperidade económica e na estabilidade social, pois representam mais de 98% dos negócios em todo mundo [4]. No mercado global atual são simultaneamente o componente mais dinâmico e vulnerável da economia [4]. Em particular, na União Europeia (UE), as PME representam 99% dos negócios e são consideradas pela Comissão Europeia (CE) a “espinha dorsal” da economia. Nos últimos 5 anos foram responsáveis por cerca de 85% dos novos empregos criados, e por dois terços do emprego no setor privado, dentro do espaço Europeu. São por isso consideradas pela CE um elemento chave para assegurar o crescimento económico, inovação, criação de emprego e integração social na UE [5].

Apesar da sua grande importância a nível socioeconómico, muitas PME, por falta de informação, indisponibilidade financeira e processos de gestão pouco inovadores, não têm uma estratégia definida em Tecnologias da Informação (TI) [3], ou seja, não utilizam plataformas de BI. Num cenário empresarial competitivo como o atual, e em particular no caso das PME, a existência de um sistema de BI é cada vez mais um ativo indispensável, uma vez que permite auxiliar na contenção de custos e otimização de recursos.

Tendo em conta a realidade de muitas PME, (reduzidos lucros e baixo número de profissionais [3]), este trabalho centra-se

nas soluções *open source* de BI, uma vez que estas, apesar de necessitarem de algum esforço na sua instalação, não têm custos com aquisição e licenças, o que as torna a opção mais viável para estas empresas. No entanto é fundamental que a PME seja capaz de escolher a solução *open source* mais adequada à sua realidade

Este artigo pretende auxiliar as PME na escolha de uma plataforma *open source* de BI. Para isso, foram selecionadas três plataformas entre as mais populares e citadas em artigos recentes [6], [7]: Jaspersoft, Knowage (SpagoBI) e Pentaho, sendo posteriormente realizada uma avaliação comparativa usando os critérios da Gartner.

Este artigo está organizado da seguinte forma. Na secção II é efetuada uma breve revisão da literatura. Na secção III são descritas as plataformas estudadas, abordando as principais características de cada uma. Na secção IV são abordados os critérios de avaliação das plataformas. Na secção V é feita a comparação entre as plataformas e, finalmente, na secção VI são apresentadas as conclusões e propostas de trabalho futuro.

## II. REVISÃO DE LITERATURA

Thomsen e Pedersen [8] foram dos primeiros autores a realizar um estudo científico sobre *open source BI*. No entanto, os autores não consideraram plataformas de BI, focando-se apenas nas funcionalidades disponíveis em ferramentas individuais. Neste trabalho, pelo contrário, são consideradas plataformas de BI *open source*, uma vez que, para uma PME, a implementação e utilização de uma plataforma adequada com as funcionalidades/critérios necessários é mais vantajosa do que a implementação de diferentes ferramentas individualmente.

Ahishakiye et al. [9] efetuaram um estudo onde realizaram uma análise comparativa de cinco plataformas *open source BI*: Apache Hadoop, Jaspersoft, Pentaho, SpagoBI e Vanilla. As funcionalidades avaliadas foram: relatórios, *Dashboards*, OLAP (*Online Analytical Processing*), ETL (*Extract, Transform and Load*), *Data Mining*, KPIs (*Key Performance Indicators*), GEO/GIS (*Geographic Information System*), *Ad-Hoc Queries*, compatibilidade com múltiplos sistemas operativos, Java, distribuição de armazenamento e processamento, tolerância ao erro e escalabilidade. Concluíram que a plataforma Apache Hadoop seria a plataforma recomendada pois era a que apresentava mais funcionalidades que eram considerados importantes, seguida pela Pentaho e a SpagoBI (também com um elevado número de critérios verificados) e, depois, a Vanilla e, por fim, a Jaspersoft. Apesar de também se focar em plataformas *open source BI*, o estudo deu maior ênfase a critérios essenciais na análise de dados criminais, o que pode ser redutor se necessitarmos de outro tipo ou de critérios adicionais para outras áreas. Também se verifica que os critérios avaliados na sua grande maioria não fazem parte dos critérios utilizados mais recentemente pela Gartner e presentes em vários estudos publicados nos anos mais recentes.

Parra et al. em [10] realizaram uma análise experimental onde compararam duas das mais bem posicionadas plataformas *open source BI* no mercado: Pentaho e Jaspersoft. O objetivo desse trabalho foi analisar e avaliar essas plataformas e descrever como melhoram a qualidade dos dados, medindo o

seu desempenho em termos de processos de ETL e *Reporting*, usando os sistemas Sage[11] e Matlab[12]. Segundo os autores, no que diz respeito à análise de processos de ETL, a plataforma Jaspersoft apresenta um incremento no tempo de CPU comparado com a Pentaho. Por outro lado, a plataforma Pentaho tem um incremento de tempo de CPU no processo de *Reporting*, comparativamente com a Jaspersoft. Sendo uma análise pormenorizada a estes dois critérios, poderá ajudar na decisão de algumas empresas para as quais estes dois critérios sejam os mais importantes, e que já se encontrem em posição de decidir entre estas duas plataformas. No entanto, não as ajuda, de uma forma geral, a perceber quais os critérios presentes em cada plataforma no seu global, isto é, qual a plataforma mais adequada às necessidades da empresa.

Golfarelli realizou uma avaliação comparativa de três plataformas: Jaspersoft, Pentaho e SpagoBI, comparando-as não só entre si, na sua versão *open source*, como também com as suas versões comerciais [13]. Concluiu que as funcionalidades administrativas poderiam ser melhoradas nas versões *open source*, sobretudo para a plataforma Pentaho e que a plataforma SpagoBI é a que oferece mais funcionalidades sendo que a sua versão *open source* é comparável às versões comerciais da Pentaho e Jaspersoft.

Lapa et al. em [3], efetuaram uma análise a 7 plataformas *open source* de BI (Actuate, Jaspersoft, OpenI, Palo, Pentaho, SpagoBI e Vanilla) segundo os critérios de Gartner 2013, aos quais considerou relevante incluir *Cloud* e *Collaborative Technologies*, dada a sua crescente importância junto dos grupos decisores. Verificou que as plataformas Pentaho e Vanilla validavam 10 em 11 dos critérios avaliados e que a SpagoBI validava todos, uma vez que era a única, à data, a disponibilizar tecnologias colaborativas. Conclui que estas 3 (Pentaho, Vanilla e SpagoBI) são competitivas com plataformas comerciais e que, embora não fosse possível identificar a melhor, tinham grande potencial de uso em PME.

Ferreira et al. [6] avaliaram 4 plataformas *open source* de BI (Birt, Jaspersoft, Pentaho e SpagoBI) utilizando a metodologia de avaliação de *software open source OSSpal*. Aplicando a metodologia, na sua escala de avaliação (de 1 a 5), a Pentaho obteve 3.47, a SpagoBI, 2.92 e a Jaspersoft, 2.90. Por comparação com a Pentaho, a SpagoBI obteve pior desempenho na categoria de comunidade e a Jaspersoft na categoria de funcionalidade.

Em [14] foram comparadas 5 plataformas *open source* de BI (Jaspersoft, Palo, Pentaho, SpagoBI e Vanilla) utilizando também os critérios da Gartner 2013. Nesta comparação, destacaram-se as plataformas Pentaho e SpagoBI, que foram submetidas a uma avaliação experimental com recurso à metodologia de comparação de *software open source*, OpenBRR. Os autores concluíram que, na escala de avaliação desta metodologia, a plataforma SpagoBI obtinha um resultado classificado como “Muito Bom”.

Apesar de algumas das plataformas abordadas nos estudos mencionados anteriormente serem as mesmas que iremos avaliar no presente artigo, os critérios avaliados que foram objeto de destaque nestes estudos não coincidem com os mais atuais utilizados pela Gartner em 2017, que serão os utilizados na nossa avaliação.

### III. PLATAFORMAS DE BI OPEN SOURCE

A seleção das plataformas a integrar neste estudo foi elaborada, numa primeira fase, com base em seis estudos comparativos que integravam plataformas *open source* BI [3], [6], [7], [14]–[16]. Foram selecionadas as 6 plataformas mais citadas. Sobre este conjunto, foram identificadas as plataformas cujos projetos *open source* continuam atualmente ativos e em desenvolvimento. Assim, as plataformas selecionadas foram: Jaspersoft, Knowage (SpagoBI) e Pentaho.

#### A. Jaspersoft

A Jaspersoft foi criada em 2004 com as ferramentas *Jasper Reports* e *iReport Designer* [17]. Posteriormente foram adicionadas a *JasperReport Server* e a *Jaspersoft OLAP* que em conjunto com as anteriores dão origem, em 2006, à Jaspersoft Business Intelligence Suite. Em 2014 a Jaspersoft foi adquirida pela TIBCO *Software*.

Esta plataforma é totalmente desenvolvida em Java e suporta Windows, Linux, Mac OS X, Apple iOS e Android. Conta atualmente com cinco versões: *Community*, *Reporting*, *AWS (Amazon Web Services)*, *Professional* e *Enterprise*. A TIBCO Jaspersoft assume um modelo de negócio *Open Source* comercial, sendo que das cinco versões apenas a *Community* é disponibilizada em licença *Open Source* (AGPL).

A Figura 1 mostra a arquitetura da versão *Community* da Jaspersoft. Conforme ilustrado na figura, a plataforma disponibiliza atualmente cinco componentes: *Jaspersoft Reports Server*, *JasperReports Library*, *Jaspersoft Studio*, *iReport Designer*, *Jaspersoft OLAP* e *Jaspersoft ETL*.



Figura 1. Arquitetura da Plataforma Jaspersoft (fonte [18])

O formato modular da instalação permite uma construção da plataforma “à medida” do projeto de BI, o que pode ser visto como uma vantagem por um lado, ou como uma desvantagem para quem pretende instalar todos os módulos já que face a soluções de instalação única é mais demorada. Esta ferramenta (na versão *Community*) apresenta três lacunas: ausência de um módulo de *Data Mining*, funcionalidades de KPIs e *Dashboards*.

Esta versão está ainda limitada face às versões comerciais na obtenção de análises *In-Memory* e embora permita interatividade nos relatórios, não é possível guardar o resultado gerado com essa interação (*write back*).

A documentação de suporte é extensa e completa, tendo esta ferramenta uma comunidade ativa.

A Figura 2 mostra um exemplo da interface da JasperReports Server.

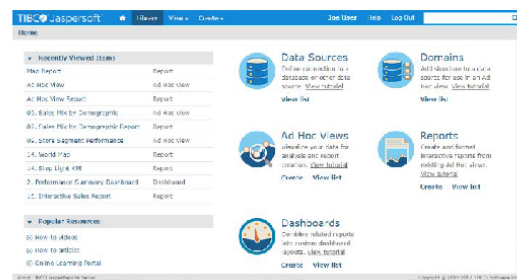


Figura 2. Interface Jaspersoft Reports Server (fonte [19])

#### B. Pentaho

A Pentaho foi criada em 2004, composta pelas ferramentas *Pentaho Reporting*, *Pentaho Reporting Server*, *Mondrian OLAP Server* e *Pentaho Data Integration* [20]. Estas ferramentas constituíam a *Pentaho Open BI Suite*. Em 2006 a Pentaho engloba os projetos *Kettle* e *Weka*. Em 2015 a Pentaho foi adquirida pela Hitachi Data Systems. Nos últimos anos tem sido lançada uma nova edição por ano, estando atualmente na versão 8.0.

A plataforma de BI da Pentaho está disponível em duas versões, ambas desenvolvidas em Java. A *Enterprise Edition*, sendo esta a comercial e a *Community Edition*, a versão *open source*. A plataforma integra os módulos *Business Analytics Platform*, *Data Integration*, *Report Designer* (Figura 3), *Aggregation Designer*, *Schema Workbench* e *Metadata Editor*.

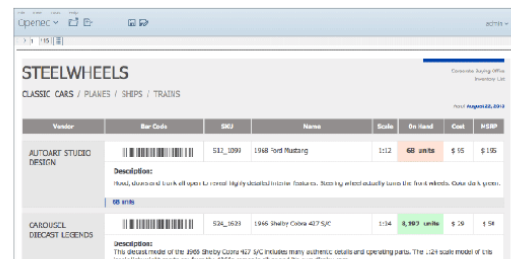


Figura 3. Exemplo de relatório da plataforma Pentaho (fonte [21])

A Pentaho destaca funcionalidades que estão presentes apenas na versão comercial. Entre elas, temos os relatórios interativos, *Ad-hoc queries*, *Drill down* e *Drill through*, *GEO/GIS*, *Dashboards* e aplicação *mobile*. Destacam ainda opções mais avançadas na integração de dados e mais aplicações em *Big Data*.

É, no entanto, possível a implementação de *Dashboards* com recurso às *Community Tools* e efetuar *Data Mining* instalando o módulo *Weka*.

A Figura 4 mostra a arquitetura da plataforma Pentaho.

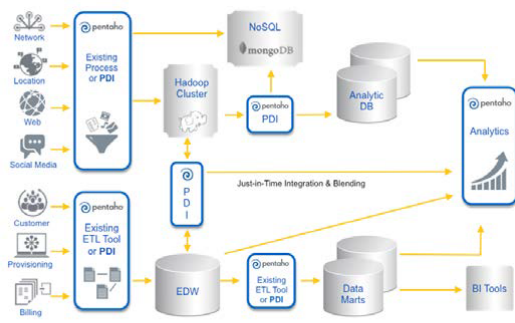


Figura 4. Arquitetura da plataforma Pentaho (fonte [22])

Tal como a JasperSoft, o formato modular da arquitetura e instalação permite uma construção da ferramenta plataforma “à medida” do projeto, o que pode ser visto como uma vantagem por um lado, ou como uma desvantagem para quem pretende instalar todos os módulos já que face a soluções de instalação única é mais demorada.

A documentação de suporte é extensa, incluindo uma plataforma *wiki*, e uma comunidade bastante ativa.

### C. Knowage (SpagoBI)

Em 2004, a SpagoWorld, uma iniciativa *open source*, fundada pela Engineering Group [23], desenvolveu, em Java, a plataforma SpagoBI. Desde junho de 2017, aquando do lançamento da versão 6.0, que a plataforma SpagoBI assumiu a denominação de Knowage [24]. A partir desse momento passou também a existir uma versão comercial (*Enterprise Edition*) e uma versão *open source* (*Community Edition*) sob licença AGPL v3 deixando de ser 100% *open source*.

A Knowage *Community Edition* (CE) mantém todas as funcionalidades da SpagoBI: *Reports*, OLAP, Gráficos, KPIs, *Dashboards* interativos, GEO/GIS, *Data Mining*, ETL, integração com MS Office e integração *mobile*.

A plataforma é composta pelos módulos *Big Data*, *Smart Intelligence*, *Enterprise Reporting* (Figura 5), *Location Intelligence*, *Performance Management* e *Predictive Analysis* que segundo a Knowage, permitem melhor escalabilidade.

A Figura 5 mostra um relatório da plataforma Knowage.

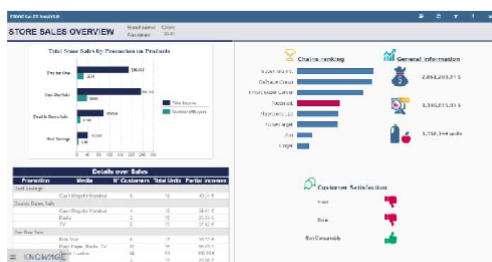


Figura 5. Exemplo de relatório da plataforma Knowage (fonte:[25])

A versão comercial contém os mesmos módulos e funcionalidades que a versão *open source*, acrescentando funções avançadas em quase todas as funcionalidades. São exemplo disso, gráficos mais interativos, nos quais é possível efetuar zoom, cockpits com atualização quase em tempo real, *what-if* com acesso a MOLAP, *self-service KPI*. Ao nível administrativo só a versão comercial permite instalação multi-ambiente, gestor de cache e gestão multi-pessoal.

A plataforma Knowage apresenta-se como uma solução de instalação única. Apenas com uma instalação no servidor, a plataforma fica pronta a operar através do *browser*. Além da versão única, a Knowage disponibiliza os módulos de forma independente, o que a torna bastante versátil no processo de instalação.

A Figura 6 mostra a arquitetura da plataforma Knowage (SpagoBI).

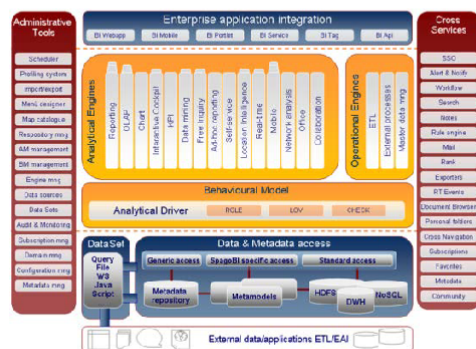


Figura 6. Arquitetura plataforma Knowage (SpagoBI) (fonte [26])

A *Community Edition* está bastante completa e a instalação *All-in-One*, acompanhada de um manual extenso e completo, é um ponto forte desta plataforma.

## IV. CRITÉRIOS DE AVALIAÇÃO

Nesta secção são descritos os critérios utilizados no nosso estudo. No presente processo de avaliação comparativa das plataformas de BI foram considerados os critérios definidos pela Gartner no Quadrante Mágico 2017 [27].

A Gartner categoriza estes critérios em 5 categorias: Infraestrutura, Gestão de Dados, Criação de Análises e Conteúdo, Partilha de Informação e Capacidades Gerais. Dos 15 critérios, foram selecionados 11: *Cloud BI*, OLAP, Fontes *Hadoop/NoSQL*, ETL, *Self-Service BI*, Informação em Tempo Real, *Dashboards*, Visualização Interativa, *Mobile BI*, Colaboração e Instalação “*All-in-One*”, uma vez que são os que apresentam maior objetividade e facilidade de avaliação. Por outro lado, por apresentarem uma avaliação mais subjetiva e de maior dificuldade, foram excluídos 4 critérios: Administração e Segurança, Gestão de Metadados, *Smart Data Discovery* e Facilidade de Uso.

Os critérios selecionados podem ser definidos da seguinte forma:

- *Cloud BI* – A plataforma de Business Intelligence é acessada através de um *browser*, estando alojada em servidores locais ou externos (Internet).

- *OLAP* – Permite que os utilizadores analisem dados com pesquisas bastante rápidas e performance de cálculo, disponibilizando um estilo de análise conhecido como “*slicing and dicing*”.

- *Fontes Hadoop/NoSQL* – Capacidade para se ligar a fontes Hadoop e NoSQL (BigData).

- *ETL* – Permite efetuar a extração de dados de diversos sistemas e fontes, transformar esses dados segundo os modelos definidos, e efetuar o seu carregamento para *Datamarts* ou *Data Warehouses*.

- *Self-Service BI* – Permite que os utilizadores façam as suas perguntas de dados, sem necessitar de um profissional de TI para criar um relatório.

- *Informação em Tempo Real* – Capacidade de devolver informação processada em tempo real nos diversos meios (*Dashboards*, *reports*, etc.)

- *Dashboards* – Capacidade de publicar, baseado em Web ou dispositivos móveis, relatórios com mostradores interativos de informação. Estes mostradores indicam o estado das métricas de desempenho comparadas com um objetivo.

- *Visualização interativa* – Oferece ao utilizador a possibilidade de visualizar numerosas vistas dos dados mais eficientemente, usando imagens e gráficos, em vez de linhas e colunas.

- *Mobile BI* – Esta capacidade permite que as organizações apresentem relatórios e *Dashboards* nas plataformas móveis, numa publicação e/ou modo interativo, aproveitando o modo de interação dos dispositivos e outras capacidades não disponíveis normalmente nos monitores.

- *Colaboração* – Permite que os utilizadores de BI partilhem e discutam a informação disponibilizada, através de *chats* e *notas*, e que possam integrar noutras ferramentas sociais ou colaborativas.

- *Instalação “All-in-One”* – Disponibiliza a instalação completa da plataforma através de um único ato de instalação por comparação a um processo de instalação por módulos.

Estes critérios são aplicados na comparação das plataformas que é apresentada na secção seguinte.

#### V. COMPARAÇÃO DE PLATAFORMAS

Cada plataforma, na sua versão *open source*, foi avaliada em cada critério. A Tabela 1 resume a comparação das plataformas, segundo os critérios definidos, assinalando-se a evidência de cada um dos critérios associados às plataformas consideradas (usando o símbolo ⊙).

Nesta comparação destaca-se, pelo número de critérios que cumpre, a plataforma Knowage, que valida 10 dos 11 critérios. A Jaspersoft e a Pentaho cumprem, 6 dos 11 critérios.

Atendendo a que são praticamente os mesmos, pode considerar-se que as duas plataformas são muito equivalentes nesta avaliação.

TABELA 1. COMPARAÇÃO PLATAFORMAS OPEN SOURCE BI

Critérios	Plataformas Open Source de BI		
	Jaspersoft	Knowage	Pentaho
Cloud BI	⊙	⊙	⊙
OLAP	⊙	⊙	⊙
Hadoop/NoSQL		⊙	⊙
ETL	⊙	⊙	⊙
Self-Service BI		⊙	
Informação em Tempo Real	⊙	⊙	⊙
Dashboards	⊙	⊙	⊙
Visualização Interativa		⊙	
Mobile BI	⊙	⊙	
Colaboração			
Instalação “All-in-One”		⊙	
<b>Total</b>	<b>6</b>	<b>10</b>	<b>6</b>

As 3 plataformas cumprem, em simultâneo, os critérios *Cloud BI*, *OLAP*, *ETL*, *Informação em Tempo Real* e *Dashboards*. Estes são critérios sobre funcionalidades que, não sendo recentes, são relevantes para a estrutura de uma plataforma de BI e é, por isso, natural a sua presença nas 3 plataformas. Ainda assim, a obtenção de *Dashboards* na Pentaho só é possível de forma livre com recurso às *Community Tools* desenvolvidas pela comunidade *open source* da Pentaho. Em *Mobile BI*, só a Pentaho tem esta funcionalidade reservada à plataforma comercial.

No âmbito da ligação a fontes *BigData*, no critério *Hadoop/NoSQL*, a Jaspersoft só disponibiliza este tipo de ligação na versão comercial da plataforma. Sobre a disponibilidade de *Self-Service BI*, a Knowage destaca-se, sendo a única a disponibilizar em *open source* esta funcionalidade. Tendo em conta a importância atual nas organizações do *Self-Service BI*, este é um ponto forte a registar na Knowage. Do mesmo modo, enquanto a Jaspersoft e a Pentaho só disponibilizam visualizações interativas na versão comercial, a Knowage disponibiliza visualizações interativas mais avançadas na sua versão *open source*.

Pela definição de colaboração da Gartner, nenhuma das plataformas verifica este critério bem como apenas a Knowage apresenta referências à colaboração, na sua versão comercial.

Um aspeto importante das plataformas é o processo de instalação onde apenas a Knowage apresenta a possibilidade de instalação “All-in-One”, sendo este um ponto forte. Esta funcionalidade reduz a complexidade do processo e também a possibilidade de falhas na integração entre módulos.

VI. CONCLUSÕES E TRABALHO FUTURO

Nesta avaliação comparativa de plataformas *open source* de BI foram incluídas as plataformas cujos projetos continuam atualmente ativos e em desenvolvimento: Jaspersoft, Knowage e Pentaho. Estas plataformas foram comparadas segundo os critérios atuais da Gartner (2017). Estes critérios foram revistos em 2016, tendo em conta mais de uma década de evolução na área do Business Intelligence e, no melhor do nosso conhecimento, esta é a primeira avaliação de plataformas *open source* de BI segundo estes critérios.

A avaliação foi efetuada com base na informação disponibilizada por cada plataforma no seu website e comunidades *open source*.

Com a aplicação dos novos critérios da Gartner ficam visíveis diferenças que, aplicando critérios antigos, não seriam detetáveis. No que diz respeito a novas tecnologias e funcionalidades, a Jaspersoft e a Pentaho relegam-nas em exclusivo para a sua versão comercial, sendo por isso as plataformas com a maior diferença entre as versões comercial e *open source*. No lado oposto temos a Knowage que apresenta uma versão *open source* muito idêntica à versão comercial.

De acordo com o nosso estudo a Knowage é a plataforma mais completa das três analisadas, destacando-se por disponibilizar uma versão de instalação "All-in-One" que agiliza todo o processo. É por isso, segundo esta comparação, a plataforma recomendada para utilização numa PME.

Esta comparação de plataformas *open source* de BI é essencialmente sobre o número de critérios cumpridos e não toma em consideração, por exemplo, a relevância/peso que cada critério/funcionalidade possa ter para uma PME. Assim, como trabalho futuro, pretendemos testar as plataformas numa PME e efetuar uma avaliação com recurso a uma metodologia que pondere critérios quantitativos e qualitativos.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] D. J. Power, "A Brief History of Decision Support Systems," *DSSResources.COM*, 2007. [Online]. Available: <http://dsresources.com/history/dshistory.html>.
- [2] T. H. Davenport, "Competing on Analytics," *Harvard Business Review*, vol. 84, no. Jan., Harvard Business Review, pp. 98–107, 2006.
- [3] J. Lapa, J. Bernardino, and A. Figueiredo, "A comparative analysis of open source business intelligence platforms," in *Proceedings of the ISDOC '14*, 2014, pp. 86–92.
- [4] Global Alliance of SMEs, "Global SME Expo 2010," 2010. [Online]. Available: <http://www.globalsmes.org/html/index.php?func=expo2010&lan=en>. [Accessed: 18-Jan-2018].
- [5] Comissão Europeia, "Entrepreneurship and Small and medium-sized enterprises (SMEs) - European Commission." [Online]. Available: <http://ec.europa.eu/growth/smes/>. [Accessed: 26-Jan-2018].
- [6] T. Ferreira, I. Pedrosa, and J. Bernardino, "Evaluating Open Source Business Intelligence Tools using OSSpal Methodology," in *Proceedings of the 9th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, 2017, pp. 283–288.
- [7] A. Brandão *et al.*, "A benchmarking analysis of open-source business intelligence tools in healthcare environments," *Inf. — Open Access J. Inf. Sci.*, vol. 7, no. 4, pp. 1–16, 2016.
- [8] C. Thomsen and T. B. Pedersen, "A Survey of Open Source Tools for Business Intelligence," *Int. J. Data Warehous. Min.*, vol. 5, no. 3, pp. 56–75, 2008.
- [9] E. Ahishakiye, E. O. Omulo, D. Taremwa, and R. Wario, "Comparative Analysis of Open source Business Intelligence tools for Crime Data Analytics," *Int. J. Latest Res. Eng. Technol.*, vol. 3, no. 4, pp. 60–65, 2017.
- [10] V. M. Parra and A. Syed, "Pentaho and Jaspersoft: A Comparative Study of Business Intelligence Open Source Tools Processing Big Data to Evaluate Performance," no. November, 2016.
- [11] "SAGE." [Online]. Available: <http://www.sage.pt/>. [Accessed: 23-Mar-2018].
- [12] "Matlab." [Online]. Available: <https://www.mathworks.com/>. [Accessed: 23-Mar-2018].
- [13] M. Golfarelli, "Open source bi platforms: A functional and architectural comparison," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 5691 LNCS, pp. 287–297, 2009.
- [14] A. Marinheiro and J. Bernardino, "Experimental Evaluation of Open Source Business Intelligence Suites using OpenBRR," *IEEE Lat. Am. Trans.*, vol. 13, no. 3, pp. 810–817, 2015.
- [15] A. Marinheiro and J. Bernardino, "OpenBRR evaluation of an open source BI suite," *Proc. Int. C\* Conf. Comput. Sci. Softw. Eng. - C3S2E '13*, vol. 1, p. 134, 2013.
- [16] M. Tereso and J. Bernardino, "Open source business intelligence tools for SMEs," in *6th Iberian Conference on Information Systems and Technologies (CISTI)*, 2011, pp. 1–4.
- [17] "Jaspersoft." [Online]. Available: <https://www.jaspersoft.com/>. [Accessed: 24-Jan-2018].
- [18] "Arquitetura Jaspersoft." [Online]. Available: <https://www.jaspersoft.com/es/jaspersoft-inicio-rapido-guia>. [Accessed: 16-Feb-2018].
- [19] "Interface Jaspersoft Reports." [Online]. Available: <https://mobiledemo.jaspersoft.com/jasperserver-pro/login.html>. [Accessed: 16-Feb-2018].
- [20] "Pentaho." [Online]. Available: <http://www.pentaho.com/>. [Accessed: 24-Jan-2018].
- [21] "Pentaho Report." [Online]. Available: <http://www.pentaho.com/product/business-analytics>. [Accessed: 16-Feb-2018].
- [22] "Arquitetura Pentaho." [Online]. Available: <https://forums.pentaho.com/threads/149369-Pentaho-5-0-blends-right-in!> [Accessed: 16-Feb-2018].
- [23] "Engeneering Group." [Online]. Available: <http://www.eng.it/>. [Accessed: 29-Jan-2018].
- [24] "Knowage." [Online]. Available: <https://www.knowage-suite.com/site/home/>. [Accessed: 24-Jan-2018].
- [25] "Relatório Knowage." [Online]. Available: <https://www.knowage-suite.com/site/product/enterprise-reporting/>. [Accessed: 16-Feb-2018].
- [26] "Arquitetura SpagoBI." [Online]. Available: [http://wiki.spagobi.org/xwiki/bin/view/spagobi\\_server/analytical\\_angular](http://wiki.spagobi.org/xwiki/bin/view/spagobi_server/analytical_angular). [Accessed: 16-Feb-2018].
- [27] J. T. Rita L. Sallam, Cindi Howson, Carlie J. Idoine, Thomas W. Oestreich, James Laurence Richardson, "Magic quadrant for business intelligence and analytics platforms," *Gartner*, no. February 2017, pp. 1–126, 2017.



## **ANEXO 2**

### **Open Source Business Intelligence Platforms' Assessment using OSSpal Methodology – ICE-B 2018**



## **Open Source Business Intelligence Platforms' Assessment using OSSpal Methodology**

Nuno Leite<sup>1</sup>, Isabel Pedrosa<sup>1</sup> and Jorge Bernardino<sup>2</sup>

<sup>1</sup> *Polytechnic of Coimbra, Coimbra Business School - ISCAC, Quinta Agrícola – Bencanta, 3040-316 Coimbra, Portugal*

<sup>2</sup> *Polytechnic of Coimbra, Institute of Engineering of Coimbra – ISEC, Rua Pedro Nunes, 3030-199 Coimbra, Portugal  
iscac16152@alumni.iscac.pt, ipedrosa@iscac.pt, jorge@isec.pt*

**Keywords:** Business Intelligence, Open Source, OSSpal Methodology, Knowage, Pentaho.

**Abstract:** The knowledge that can be acquired from existing data in organizations is critical to increasing organizations competitive advantage in today's changing markets. The use of Business Intelligence (BI) platforms is an effective choice in support of decision-making. BI platforms are a major asset for any enterprise, as they have multiple benefits, such as efficient use of resources, identification of business opportunities and negative trends that become a competitive advantage. Open source BI platforms provide most of the functionalities available in commercial solutions without increasing costs for enterprises. However, it is important to know which open source BI platform to choose. In this paper, it is used OSSpal, an open source assessment methodology to evaluate two of the most popular open source BI platforms: Knowage and Pentaho.

### **1 INTRODUCTION**

Business Intelligence translates into a set of management practices, implemented through software, with the objective of increasing profitability and supporting administrations in the decision-making and leadership of their organizations (Lapa, Bernardino and Figueiredo, 2014). Enterprises that use BI platforms have analytical tools that provide important information and data for their management.

The term Business Intelligence (BI) was introduced by Howard Dresner of the Gartner Group in 1989 (Power, 2007). Davenport defines a BI platform as a set of processes and software used to collect, analyse and disseminate data, with the aim of better decision making (Davenport, 2006). BI platforms use data available in organizations to generate and deliver information used to support decision-making. This information is obtained by combining data interrogation and exploration tools with tools that enable reporting. These platforms typically associate three technologies: Data Warehouses, On-Line Analytical Processing (OLAP) and Data Mining. Data Warehouse is an integrated repository that allows storing information. This information can then be analysed through OLAP and / or Data Mining tools.

OLAP is a multidimensional analysis that allows analysing the information under different perspectives. Data Mining uses data mining algorithms that identify patterns, relationships, models, etc. Business Intelligence contributes to increase the collective intelligence, learning ability and creativity of the organization (Santos and Ramos, 2006). This work focuses on open source BI platforms. Although they require some effort in their installation, they have no acquisition costs and licenses, which makes them the most viable option for enterprises (Lapa, Bernardino and Figueiredo, 2014).

The increase in the use of Open Source Software in its "Free / Libre" Open Source Software (FOSS) aspect that we witness at the beginning of the 21st century is due to several factors, including the absence of licensing costs and the availability of source code that allows users to tailor it to their specific needs. A disadvantage is the absence of metrics that assure the quality of this and prove its validity (Petrinja, Sillitti and Succi, 2008).

It becomes fundamental that the enterprises make an informed choice regarding open source software. In order to assist enterprises in this task, and to address this main objective, in this paper we apply the OSSpal methodology to assess two open source BI platforms: Knowage and Pentaho. To the best of our knowledge, it is the first time Knowage is assessed

with OSSpal methodology and the installation process is carried out to provide a better assessment.

The remainder of this paper is organized as follows. Section 2 presents the related work. Section 3 describes the two open source BI platforms assessed. Section 4 explains the fundamentals of OSSpal methodology and Section 5 presents the assessment of the platforms with OSSpal. Finally, Section 6 presents the conclusions and future work.

## **2 RELATED WORK**

The advent of FLOSS made the traditional software evaluation models like McCall, Boehm's or ISO 9126, not applicable to all software. These models cannot be adapted to the Open Source development practices and thus, cannot be used to evaluate the software and its community as a whole (Samoladas and Gousios, 2008).

Deprez and Alexandre (2008) conducted the first effort comparing FLOSS assessment methodologies. They have done a rigorous comparison between Open Business Readiness Rating (OpenBRR) and Qualification and Selection of Open Source Software (QSOS) based on the description of the methodologies and not on their empirical application. They identified advantages and disadvantages of both methodologies. They concluded that OpenBRR allows tailoring the criteria to a domain, hence a better fit to the evaluation context, but terminology is broad and imprecise for the top nodes in the hierarchy. On the other hand, QSOS has an extensive list of criteria but the scoring rules are ambiguous for more than half of the criteria. The authors also conclude that QSOS 3-level score is too restrictive.

Petrinja *et al.* (2010) developed a study on the quality and usability of three FLOSS assessment models: OpenBRR, the QSOS, and the QualiPSO OpenSource Maturity Model (OMM). The study identified the positive and negative aspects of each of them. The results revealed that the three models provided comparable assessments. The main conclusion was that all three models have some questions that do not have a clear formulation and thus are not clear to the assessors. In some questions, the threshold value available for the answer was not clear either. The critical aspects of each model were: Functionality and Quality for OpenBRR; Adoption, Administration/Monitoring, Copyright owners, and Browser for QSOS; and Quality of the Test Plan, and the Technical Environment for OMM.

In Marinheiro and Bernardino (2015), five open source BI platforms (Jaspersoft, Palo, Pentaho,

SpagoBI and Vanilla) were compared using Gartner 2013 criteria. In this comparison, they highlight the Pentaho and SpagoBI platforms, which were submitted to an experimental evaluation using the methodology of open source software comparison, OpenBRR. The authors concluded that, in the evaluation scale of this methodology, the SpagoBI platform obtained the best result.

Ferreira, Pedrosa, and Bernardino (2017) evaluated four open source BI platforms (Birt, Jaspersoft, Pentaho and SpagoBI) using the OSSpal methodology. Applying the methodology, in its scale of evaluation (from 1 to 5), Pentaho obtained 3.47, SpagoBI 2.92 and Jaspersoft 2.90. Compared to Pentaho, SpagoBI performed poorly in the community category and Jaspersoft in the functionality category.

Leite, Pedrosa, and Bernardino (2018) developed a comparative evaluation of three open source BI platforms (Jaspersoft, Knowage and Pentaho) using Gartner's 2017 criteria. According to the authors, Knowage is the new version of SpagoBI that now has also a commercial version and no longer is 100% open source. In that evaluation, Knowage validated 10 out of 11 criteria while Jaspersoft and Pentaho validated 6 of the 11 criteria. The authors concluded that, with the new Gartner criteria, differences became clearer among these three platforms: while Knowage has almost the same main functionalities in their commercial and open sources version, Jaspersoft and Pentaho relegate the new features only to their commercial versions.

Although some of the platforms addressed in the previous research mentioned are the same ones that we will assess in this paper, in none of the studies Knowage and Pentaho BI platforms were installed and tested. In addition, to the best of our knowledge, Knowage has never been assessed with OSSpal.

## **3 BUSINESS INTELLIGENCE PLATFORMS**

In a previous comparative evaluation, we analysed three platforms using Gartner's 2017 criteria: Jaspersoft, Knowage and Pentaho. The three open source projects were identified, out of six projects, as the ones still active and under development.

Knowage obtained the best result in this evaluation while Jaspersoft and Pentaho performed equally. Based on these results, Knowage is the first BI platform selected for the assessment.

In a number of recent studies (Tereso and Bernardino, 2011; Marinheiro and Bernardino, 2015; Ferreira, Pedrosa and Bernardino, 2017) Pentaho has best scores than Jaspersoft. In addition, once compared at Google Trends, Pentaho scores 83 while Jaspersoft scores 20. Therefore, Pentaho is the second BI platform selected for this assessment.

Next, we briefly describe Knowage and Pentaho BI platforms.

### 3.1 Knowage

In 2004, SpagoWorld, an open source initiative founded by the Engineering Group, developed the SpagoBI platform in Java. Since June 2017, at the time of the release of version 6.0, the platform SpagoBI assumed the new designation of Knowage. From that moment on, two licenses became available: a commercial version (Enterprise Edition) and an open source version (Community Edition) under AGPL v3 license, ceasing to be 100% open source.

Knowage Community Edition (CE) maintains all SpagoBI features: Reports, OLAP, Graphs, KPIs, Interactive dashboards, GEO / GIS, Data Mining, MS Office integration and mobile integration.

The platform is composed of the following modules: Big Data, Smart Intelligence, Enterprise Reporting, Location Intelligence, Performance Management and Predictive Analysis. According to Knowage they allow better scalability and are described next.

- *Big Data*: allows to not only work with large volumes of data, but also combine different sources so you can develop different analyses.
- *Smart Intelligence*: enables the development of static reports, maps, interactive cockpits as well as ad-hoc queries via drag & drop and multidimensional analysis (OLAP). The CE version does not allow calculated field, time series and Multidimensional Expressions (MDX) functions at the OLAP level.
- *Enterprise Reporting*: produces reports such as the one shown in Figure 1 and allows exporting to various formats including PDF and MS Office. It also allows scheduling offline reports and distributing them to a set of selected users.
- *Location Intelligence*: is a module dedicated to the spatial analysis of information, using various types of sources such as maps or vector images (SVG). It allows working traditional information with spatial information that has a relation between them, producing dynamic maps.

- *Performance Management*: is a module dedicated to the production and visualization of KPIs and scorecards.
- *Predictive Analysis*: enables advanced processing with Data Mining techniques to simulate actions and to evaluate their effects. For the “what-if” feature, this module uses an OLAP solution that allows interactive simulation between measurements and dimensions via drag & drop.

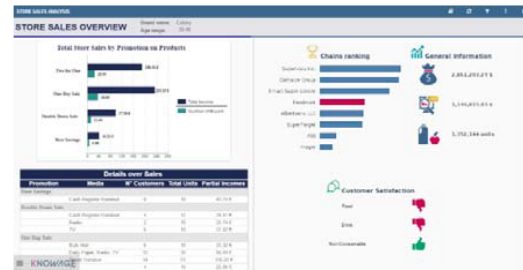


Figure 1: A report from Knowage platform (Knowage, 2018).

The commercial version contains the same modules and functionalities as the open source version, but it adds advanced functions to almost all the modules. Examples of this are more interactive graphs in which we can zoom, cockpits with near real-time updates, what-if with access to MOLAP, and self-service KPIs. At the administrative level, only the commercial version allows multi-environment installation, cache manager and multi-person management.

The Knowage platform is presented as an all-in-one installation solution. With only one installation on the server, the platform is ready to operate through the browser. In addition to the single version, Knowage provides the modules independently, which makes it quite versatile in the installation process.

The Community Edition is quite complete and the all-in-one installation, accompanied by an extensive and comprehensive manual, is a strong point of this platform.

### 3.2 Pentaho

Pentaho was created in 2004, comprising Pentaho Reporting, Pentaho Reporting Server, Mondrian OLAP Server and Pentaho Data Integration tools. These tools composed the Pentaho Open BI Suite. In 2006, Pentaho encompasses the Kettle and Weka projects. In 2015, Hitachi Data Systems acquired

Pentaho. In the last years has been released a new edition per year, being currently in version 8.0.

The Pentaho BI platform is available in two versions, both developed in Java. The Enterprise Edition, this being the commercial and the Community Edition, the open source version. The platform integrates the following modules:

- *Business Analytics Platform*: is the server that provides various services to users such as reports and integration tools.
- *Data Integration*: is the platform's ETL module, also known as Kettle, and allows data extraction, transformation and loading actions.
- *Report Designer*: is a graphical tool that allows you to design reports as shown in Figure 3.
- *Aggregation Designer*: allows to create and maintain aggregate tables.
- *Schema Workbench*: is a visual interface for creating and testing OLAP cubes in Mondrian.
- *Metadata Editor*: presents itself as a tool that simplifies the reporting experience by allowing to build metadata domains and relational data models.

Year	Bar Code	ID#	Name	Scale	On Hand	Cost	MSRP
1958	512_099	1958 Ford Mustang	1:22	488 units	\$ 55	\$ 195	
1966	524_1628	1966 Shelby Cobra 427 5/C	1:24	8,187 units	\$ 29	\$ 50	

Figure 2: An example of a report designed with Pentaho platform.

Pentaho highlights features that are only present in the commercial version. Among them, the interactive reports, Ad-hoc queries, Drill down and Drill through, GEO / GIS, Dashboards and mobile application. They also highlight more advanced options in data integration and more sources in Big Data. However, it is possible to implement Dashboards with Community Tools.

The modular format of Pentaho architecture and installation allows the users to build a platform “tailored” to their needs. This is an advantage but considering the installation consumes more time, some users may consider it a disadvantage.

The support documentation is extensive, including a help website (help.pentaho.com), and a very active community (community.hds.com).

## 4 OSSPAL METHODOLOGY

OSSpal has emerged as a successor of the Business Readiness Rating (OpenBRR) with the goal to provide a trusted, unbiased source for evaluation of open source software. It aims to be an open, comprehensive and standard assessment model that is trusted, widely used and “tunable” (Wasserman, 2014). OSSpal combines quantitative and qualitative evaluation measures to decide which software has the best score. This way it can assist companies, government agencies, and other organizations in finding high quality free open source software (Wasserman *et al.*, 2017).

The implementation of OSSpal Methodology is composed of four phases (OpenBRR, 2005):

### Phase 1: Quick Assessment Filter

Identification of the components of the software to be analysed, measuring each component in relation to the evaluation criteria.

### Phase 2: Target Usage Assessment

Allocation of weights to categories and measures:

- a. Assign a percentage of importance to each category. They should sum up 100%.
- b. For each measure within a category, rank the measure according to its importance.
- c. Assign a percentage to each measure within a category according to its importance, totalling 100% over all the measures within one category.

### Phase 3: Data collection and Processing

Gather data for each metric used in each category rating, and calculate the applied weighting for each metric, at a level of 1 (unacceptable) to 5 (excellent).

### Phase 4: Data Translation

Use category ratings and the functional orientation weighting factors to calculate the OSSpal final score.

The OSSpal methodology, shown in Figure 5, consists of seven evaluation areas (Wasserman *et al.*, 2017):

- **Functionality**: How well will the software meet the average user's requirements?
- **Operational Software Characteristics**: How secure is the software? How well does the software perform? How well does the software scale to a large environment? How good is the UI? How easy to use is the software for end-

users? How easy is the software to install, configure, deploy, and maintain?

- **Support and Service:** How well is the software component supported? Is there commercial and/or community support? Are there people and organizations that can provide training and consulting services?
- **Documentation:** Is there adequate tutorial and reference documentation for the software?
- **Software Technology Attributes:** How well is the software architected? How modular, portable, flexible, extensible, open, and easy to integrate is it? Are the design, the code, and the tests of high quality? How complete and error-free are they?
- **Community and Adoption:** How well is the component adopted by community, market, and industry? How active and lively is the community for the software?
- **Development Process:** What is the level of the professionalism of the development process and of the project organization as a whole?

Functionality is an assessment category that is computed differently from other categories. Each type of software application has a unique set of features that needs to be fulfilled by the software. The Functionality rating is obtained by first comparing the features of the component being evaluated with a standard feature-set required for an average use. This standard feature-set must be constructed, or borrowed from an external source (Phase 1).

The following steps should be used to compute de Functionality score:

- i. Assign an importance score to all items in the feature list, using a scale of 1 to 3, with 1 being less important, 3 being very important.
- ii. Compare the feature list of the component with the standard feature list. For each feature met, add the importance score to a cumulative sum. If not met, deduct importance score from the sum.
- iii. Divide the cumulative sum by the maximum score that can be obtained by the standard features. This ratio is called the feature score.
- iv. Normalize the feature score to a scale of 1 to 5 using this scheme:
  - Under 65%, score = 1 (unacceptable)
  - 65% - 80%, score = 2 (bad)
  - 80% - 90%, score = 3 (acceptable)
  - 90% - 96%, score = 4 (very good)
  - Greater than 96%, score = 5 (excellent)

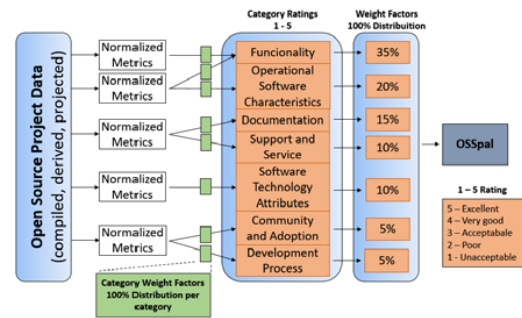


Figure 3: OSSpal methodology.

## 5 EVALUATION

OSSpal appears as the successor to OpenBRR, combining a qualitative and quantitative evaluation of the software. It aims to assist companies, government agencies, and other organizations in finding high quality FLOSS (Wasserman *et al.*, 2017). To ensure a more reliable and accurate assessment using OSSpal, the installation process was carried out for both open source BI platforms. The installation was followed by a basic use in order to provide user experience.

As stated in Phase 1, the features list was elaborated to the functionality category. We selected our feature list following the criteria used by Leite, Pedrosa and Bernardino (2018) which are based on Gartner 2017 Magic Quadrant for Business Intelligence and Analytics Platforms. These features allow a more objective assessment. With the features list elaborated, an importance score was assigned to each feature from 1 to 3 (less to very important).

Table 1 shows the features chosen for the functionality category and the weights given to each one, according to the OSSpal methodology.

Table 1: Weights assigned to each feature in the functionality category.

Features	Weight
Dashboards	3
Interactive Visualization	3
OLAP	3
Real Time Information	3
ETL	2
Mobile BI	2
Self-Service BI	2
All-in-One Installation	1
Cloud BI	1
Collaboration	1
Hadoop/NoSQL	1

As stated in Phase 2, we allocated weights for each category totalling 100%, as showed in Table 2.

Table 2: Weights assigned to each category.

Category	Weight
Functionality	35%
Operational Software Characteristics	20%
Documentation	15%
Support and Service	10%
Software Technology Attributes	10%
Community and Adoption	5%
Development Process	5%
<b>Total</b>	<b>100%</b>

We considered “Functionality” the most important category as it consists on the core of the software assessed. For this reason, it has been given the highest weight (35%). Following with a weight of 20%, we considered “Operational Software Characteristics” the second most important category as it considers into evaluation areas like user experience and installation process. Still with some importance, with weights of 15% and 10%, follows the “Documentation” and “Support and Service” categories respectively. Especially in the open source context, these categories play an important role on helping users and Information Technologies professionals. Considered of less relevance, the category of “Software Technology Attributes” was given a 10% weight, “Community and Adoption” and “Development Process” categories where both weighted 5%.

After this weight attribution to all categories, Phase 3 takes place. Each BI platform is assessed and for each category, a score from 1 (unacceptable) to 5 (excellent) is given.

As stated previously, the score from 1 to 5 for functionality category is computed differently.

Table 3 presents the intermediate results for this step and the score obtained for functionality category.

In Phase 4, all the scores are translated according to the weight each category was given (e.g., 10% of 5 translates to 0.5). The cumulated sum of each category-translated score gives the final score of each BI platform.

Table 4 presents the results of the assessment.

Pentaho, with a score of 4.35 (from 1 to 5) was the BI platform with the highest score. Knowage has scored 3.31. Pentaho scores slightly better than Knowage on each category, except Functionality where it has a difference of 0.35. In the first step to compute Functionality score, Knowage had a result of 86% and Pentaho 91%.

Table 3: Functionality score.

Feature	Weight	Knowage	Pentaho
Dashboards	3	3	3
Interactive Visualization	3	3	3
OLAP	3	3	3
Real Time Information	3	3	3
ETL	2	0	2
Mobile BI	2	2	2
Self-Service BI	2	2	2
All-in-One Installation	1	1	0
Cloud BI	1	1	1
Collaboration	1	0	0
Hadoop/NoSQL	1	1	1
<b>Cumulative sum</b>	<b>22</b>	<b>19</b>	<b>20</b>
<b>Normalization to scale 1-5</b>	<b>100%</b>	<b>86%</b>	<b>91%</b>
		<b>3</b>	<b>4</b>

While this stands for a close result, the normalization set by OSSpal methodology transforms this value in a score of 3 to Knowage and 4 to Pentaho.

Table 4: OSSpal final score.

Category	Score	
	Knowage	Pentaho
Functionality	1.05	1.40
Operational Software Characteristics	0.80	1.00
Documentation	0.53	0.68
Support and Service	0.35	0.45
Software Technology Attributes	0.30	0.40
Community and Adoption	0.13	0.23
Development process	0.15	0.20
<b>TOTAL</b>	<b>3.31</b>	<b>4.35</b>

Applying the 35% weight to these scores, means a rather relevant impact on the final score than it actually was at the beginning.

In Operational Software Characteristics, Pentaho’s user interface is simpler than Knowage’s, yet more intuitive and effective. As for Documentation, Pentaho has more and better tutorials, which is important on the FLOSS context.

The final difference between Pentaho and Knowage scores is 1.04. We address this difference with the fact that Pentaho has a much larger worldwide adoption, which helps to become a more mature software.

## 6 CONCLUSIONS AND FUTURE WORK

In this paper, we evaluated two open source BI platforms still active and under development. This evaluation was developed using OSSpal, which is an open source software assessment methodology. The use of an assessment methodology for open source software is highly recommended as it allows to achieve more reliable results.

The information required to develop the evaluation was gathered from the websites of the BI platforms. In addition, the installation process of each open source version of the platforms was made and a basic user experience of the software took place. This allowed to confirm the information gathered from the websites and to better evaluate some of the categories that make part of the OSSpal methodology. Pentaho presented the best score after applying the OSSpal methodology. Knowage scored less than Pentaho but it has the potential to perform better in the future.

Knowage has an “All-in-One” package for installation that simplify the process and the core of the platform was up and running in about half an hour. Pentaho has more steps to achieve the same stage but if all instructions are followed correctly, it can be working in less than an hour.

The overall conclusion is that Pentaho is a more mature software than Knowage in all categories and this is the result of a much larger worldwide use and community.

As future work, we intend to create measures under each assessment category and to perform a more extended used of the platforms by developing a real case study scenario.

## REFERENCES

- Davenport, T. H. (2006) ‘Competing on Analytics’, *Harvard Business Review*, pp. 98–107.
- Deprez, J. C. and Alexandre, S. (2008) ‘Comparing Assessment Methodologies for Free/Open Source Software: OpenBRR and QSOS’, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5089 LNCS, pp. 189–203. doi: 10.1007/978-3-540-69566-0\_17.
- Ferreira, T., Pedrosa, I. and Bernardino, J. (2017) ‘Evaluating Open Source Business Intelligence Tools using OSSpal Methodology’, in *Proceedings of the 9th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, pp. 283–288. doi: 10.5220/0006516402830288.
- Knowage (2018) *Knowage Report*. Available at: <https://www.knowage-suite.com/site/product/enterprise-reporting/> (Accessed: 16 February 2018).
- Lapa, J., Bernardino, J. and Figueiredo, A. (2014) ‘A comparative analysis of open source business intelligence platforms’, in *Proceedings of the ISDOC '14*, pp. 86–92. doi: 10.1145/2618168.2618182.
- Leite, N., Pedrosa, I. and Bernardino, J. (2018) ‘Comparative Evaluation of Open Source Business Intelligence Platforms for SME’, in *13th Iberian Conference on Information Systems and Technologies (CISTI)*.
- Marinheiro, A. and Bernardino, J. (2015) ‘Experimental Evaluation of Open Source Business Intelligence Suites using OpenBRR’, *IEEE Latin America Transactions*, 13(3), pp. 810–817. doi: 10.1109/TLA.2015.7069109.
- OpenBRR (2005) *OpenBRR White paper*. Available at: [https://web.archive.org/web/20050803022846/http://www.openbrr.org/docs/BRR\\_whitepaper\\_2005RFC1.pdf](https://web.archive.org/web/20050803022846/http://www.openbrr.org/docs/BRR_whitepaper_2005RFC1.pdf) (Accessed: 12 April 2018).
- Petrinja, E., Sillitti, A. and Succi, G. (2008) ‘Overview on trust in large FLOSS communities’, *IFIP International Federation for Information Processing*, 275, pp. 47–56. doi: 10.1007/978-0-387-09684-1\_4.
- Petrinja, E., Sillitti, A. and Succi, G. (2010) ‘Comparing OpenBRR, QSOS, and OMM Assessment Models’, in *IFIP International Federation for Information Processing*, pp. 224–238. doi: [https://doi.org/10.1007/978-3-642-13244-5\\_18](https://doi.org/10.1007/978-3-642-13244-5_18).
- Power, D. J. (2007) *A Brief History of Decision Support Systems*, *DSSResources.COM*. Available at: <http://dssresources.com/history/dsshistory.html>.
- Samoladas, I. and Gousios, G. (2008) ‘The SQO-OSS quality model: measurement based open source software evaluation’, *Open Source Development, Communities and Quality. OSS 2008. IFIP – The International Federation for Information Processing*, 275, pp. 237–248. doi: 10.1007/978-0-387-09684-1\_19.
- Santos, M. Y. and Ramos, I. (2006) *Business Intelligence : tecnologias da informação na gestão de conhecimento*. FCA - Editora de Informática. Available at: <https://repositorium.sdum.uminho.pt/handle/1822/6198> (Accessed: 15 July 2017).
- Tereso, M. and Bernardino, J. (2011) ‘Open source business intelligence tools for SMEs’, in *6th Iberian Conference on Information Systems and Technologies (CISTI)*, pp. 1–4. Available at: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5974187>.
- Wasserman (2014) *OSSpal Model*. Available at: [http://oss.sv.cmu.edu/presentations/WeekThree2014\\_Wasserman.pdf](http://oss.sv.cmu.edu/presentations/WeekThree2014_Wasserman.pdf) (Accessed: 12 January 2018).
- Wasserman, A. I. et al. (2017) ‘OSSpal: Finding and Evaluating Open Source Software’, in *Open Source Systems: Towards Robust Practices*, pp. 193–203. doi: 10.1007/978-3-319-57735-7.

## **ANEXO 3**

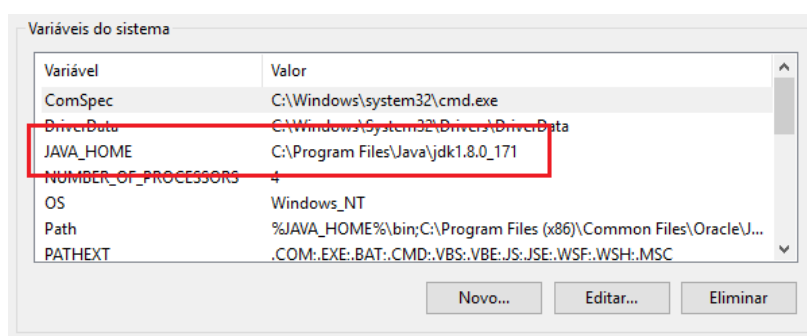
### **Preparação da Plataforma Pentaho**



## Preparação da Plataforma

### Instalação do Java

A plataforma Pentaho é desenvolvida em Java. O primeiro procedimento a executar é instalar o Java *Development Kit* (JDK) e neste caso foi instalado o JDK 8u171. De seguida é necessário criar uma variável de ambiente. No Painel de Controlo fomos a Sistema > Definições Avançadas do Sistema > Variáveis de Ambiente e criamos uma nova variável JAVA\_HOME com o caminho onde está instalado o JDK, conforme podemos ver na seguinte.



### Instalação da Pentaho

A versão *open source* da Pentaho 8.0 está disponível para *download* na plataforma SourceForge. Foi efetuado o *download* do ficheiro pentaho-server-ce-8.0.0.0-28.zip em <https://sourceforge.net/projects/pentaho>. O ficheiro deve ser descompactado por forma a que fique na pasta C:\pentaho\server\pentaho-server. Este é o elemento principal da plataforma e vem quase pronto a funcionar.

### Instalação do Pentaho Report Designer (PRD)

Para a instalação do PRD foi efetuado o *download* do ficheiro prd-ce-8.0.0.0-28.zip na plataforma SourceForge em <https://sourceforge.net/projects/pentaho>. Deve ser descompactado por forma a que fique na pasta C:\pentaho\design-tools\report-designer.

### **Instalação do Pentaho Schema Workbench (PSW)**

Para a instalação do PSW foi efetuado o *download* do ficheiro psw-ce-8.1.0.0-365.zip na plataforma SourceForge em <https://sourceforge.net/projects/pentaho>. Deve ser descompactado por forma a que fique na pasta C:\pentaho\design-tools\schema-workbench.

### **Instalação do PostgreSQL**

O Pentaho funciona com as versões 9.5 e 9.6 do PostgreSQL, sendo que a versão disponibilizada vem configurada para o PostgreSQL 9.6. Assim foi efetuado o *download* e instalação do PostgreSQL 9.6.9-1. Durante a instalação do PostgreSQL é registado um utilizador e password de administrador que deve ser guardado para utilização futura.

### **Instalação do pgAdmin4**

O pgAdmin é a ferramenta de administração de bases de dados para o PostgreSQL. Foi efetuado o *download* e instalação do pgAdmin4 v3.0.

### **Criação das Bases de Dados para o Pentaho**

O Pentaho necessita que sejam criadas três bases de dados para funcionar. São elas a jrc, quartz e repository. Para o efeito, existem três *scripts* SQL que podem ser utilizados. Neste caso em que utilizamos uma base de dados PostgreSQL, estes *scripts* podem ser encontrados na pasta C:\pentaho\server\pentaho-server\data\postgresql. Para facilitar o processo, sugerimos que estes ficheiros sejam copiados para C:\.

Para criar as bases de dados, devemos invocar os *scripts* através da SQL Shell que foi instalada junto com o pgAdmin4. Iniciada a Shell, devem premir <Enter> até ser solicitada a password. Aqui devem utilizar a password guardada anteriormente. Já na linha de comando da Shell, devem correr os seguintes comandos para correr os *scripts*:

```
“\i 'c:/create_quartz_postgresql.sql' ”  
“\i 'c:/create_repository_postgresql.sql' ”  
“\i 'c:/create_jcr_postgresql.sql' ”
```

Deixamos uma chamada de atenção ao pormenor da barra inclinada à direita e não à esquerda em ‘c:/’ para que a linha de comandos interprete corretamente o caminho indicado para a localização do script SQL. Na seguinte, exemplo de execução do script de criação da base de dados repository.

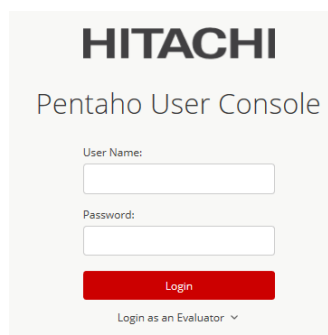
```
postgres=# \i 'c:/create_repository_postgresql.sql'
psql:c:/create_repository_postgresql.sql:7: NOTICE: database "hibernate" does not exist, skipping
DROP DATABASE
psql:c:/create_repository_postgresql.sql:8: NOTICE: role "hibuser" does not exist, skipping
DROP ROLE
CREATE ROLE
CREATE DATABASE
GRANT
postgres=#
```

### **Acesso à plataforma Pentaho**

Antes de aceder ao Pentaho é necessário iniciar o servidor, correndo a *batch* ‘C:\pentaho\server\pentaho-server\start-pentaho.bat’. Para desligar o servidor é recomendável o uso da *batch* ‘C:\pentaho\server\pentaho-server\stop-pentaho.bat’. O servidor está pronto a funcionar quando termina de correr todos os processos e informa do tempo decorrido no processo como mostra a figura seguinte.

```
02-Jun-2018 19:30:44.667 INFO [main] org.apache.coyote.AbstractP
rotocol.start Starting ProtocolHandler ["http-apr-8080"]
02-Jun-2018 19:30:44.682 INFO [main] org.apache.coyote.AbstractP
rotocol.start Starting ProtocolHandler ["ajp-apr-8009"]
02-Jun-2018 19:30:44.688 INFO [main] org.apache.catalina.startup
.Catalina.start Server startup in 63270 ms
```

O acesso à plataforma é feito em qualquer *browser* compatível, através do endereço ‘localhost:8080/pentaho’. Surge o painel de login conforme a figura seguinte. Os dados de acesso por defeito são “admin” para utilizador e “password” para a palavra passe.



**HITACHI**  
Pentaho User Console

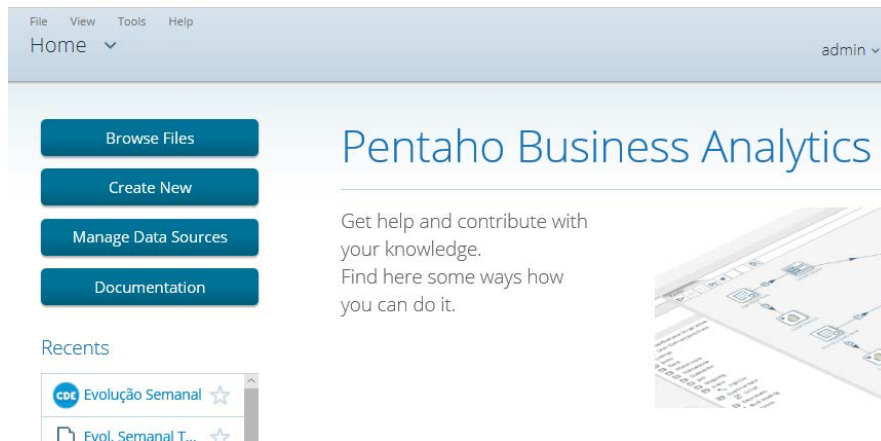
User Name:

Password:

Login

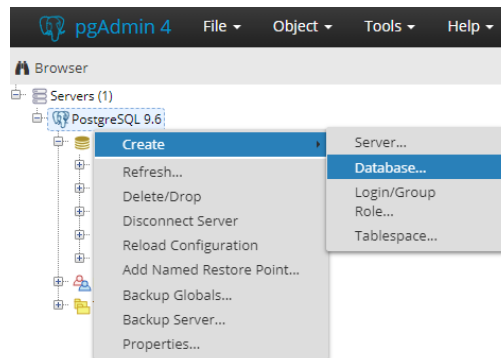
Login as an Evaluator ▾

A figura seguinte mostra o ambiente de trabalho principal do Pentaho.

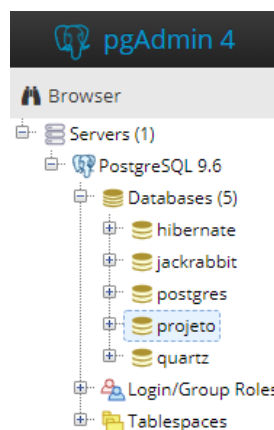


### **Criação da Base de Dados para a Good Mailman**

Na criação desta Base de Dados, utilizamos o pgAdmin4. Para o efeito, basta um click com o botão direito do rato sobre “PostgreSQL 9.6” e criar a base de dados como mostra a figura seguinte.



Na figura abaixo, podemos ver as bases de dados criadas anteriormente bem como a base de dados criada neste passo, aqui designada por “projeto”. Posteriormente iremos criar as tabelas e importar os dados.

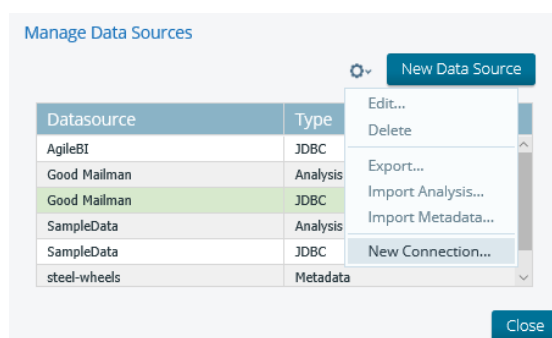


## Instalação do Plugin Saiku

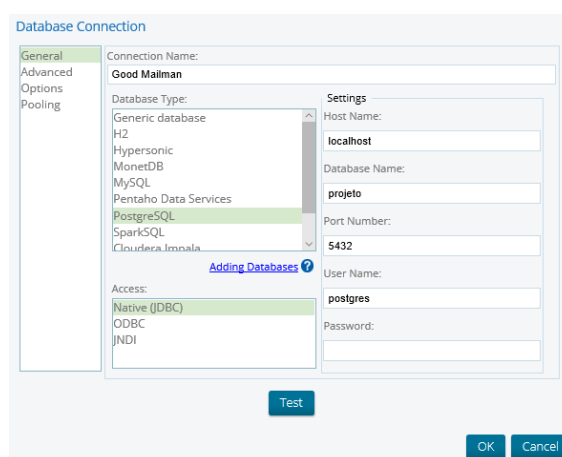
Para instalar o Saiku, dentro da plataforma Pentaho, vamos a “Home > Marketplace”. No Marketplace pesquisamos por “Saiku” e instalamos a versão *community* do “Saiku Analytics”. Teremos de nos registar e obter uma licença, mas todo o processo é gratuito.

## Ligação a fonte de dados no Pentaho

Por fim, temos de criar uma ligação à base de dados. Para tal, vamos a “Manage Data Sources > New Connection” como ilustrado na figura seguinte.



A figura abaixo ilustra o preenchimento dos dados para a ligação. O nome de utilizador e palavra passe deverão ser os registados anteriormente aquando da instalação do PostgreSQL.



## **ANEXO 4**

### **Código SQL das Tabelas criadas em PostgreSQL**

## **Código SQL das Tabelas criadas em PostgreSQL**

### **Tabela fact\_cartas**

```
CREATE TABLE public.fact_cartas
(
    id_carta integer NOT NULL,
    data_imp date,
    data_entrega date,
    cliente integer,
    centro integer,
    codigo_postal integer,
    data_imp_datekey integer,
    data_entrega_datekey integer,
    status integer,
    idade integer,
    CONSTRAINT fact_cartas_pkey PRIMARY KEY (id_carta),
    CONSTRAINT centro FOREIGN KEY (centro)
        REFERENCES public.dim_centro (id_centro) MATCH
FULL
        ON UPDATE NO ACTION
        ON DELETE NO ACTION,
    CONSTRAINT cliente FOREIGN KEY (cliente)
        REFERENCES public.dim_cliente (id_cliente)
MATCH FULL
        ON UPDATE NO ACTION
        ON DELETE NO ACTION,
    CONSTRAINT codigo_postal FOREIGN KEY
(codigo_postal)
        REFERENCES public.dim_codigo_postal
(id_codigo_postal) MATCH FULL
        ON UPDATE NO ACTION
        ON DELETE NO ACTION,
    CONSTRAINT data_entrega FOREIGN KEY
(data_entrega_datekey)
        REFERENCES public.dim_date ("DateKey") MATCH
FULL
        ON UPDATE NO ACTION
        ON DELETE NO ACTION,
    CONSTRAINT data_imp FOREIGN KEY (data_imp_datekey)
        REFERENCES public.dim_date ("DateKey") MATCH
FULL
        ON UPDATE NO ACTION
        ON DELETE NO ACTION,
    CONSTRAINT idade FOREIGN KEY (idade)
        REFERENCES public.dim_idade (id_idade) MATCH
SIMPLE
        ON UPDATE NO ACTION
        ON DELETE NO ACTION,
    CONSTRAINT status FOREIGN KEY (status)
```

```
REFERENCES public.dim_status (id_status) MATCH
FULL
ON UPDATE NO ACTION
ON DELETE NO ACTION
)
WITH (
    OIDS = FALSE
)
TABLESPACE pg_default;

ALTER TABLE public.fact_cartas
    OWNER to postgres;
```

### **Tabela dim\_centro**

```
CREATE TABLE public.dim_centro
(
    id_centro integer NOT NULL,
    nome_centro text COLLATE pg_catalog."default" NOT
NULL,
    CONSTRAINT dim_centro_pkey PRIMARY KEY (id_centro)
)
WITH (
    OIDS = FALSE
)
TABLESPACE pg_default;

ALTER TABLE public.dim_centro
    OWNER to postgres;
```

### **Tabela dim\_cliente**

```
CREATE TABLE public.dim_cliente
(
    id_cliente integer NOT NULL,
    nome_cliente character(1) COLLATE
pg_catalog."default" NOT NULL,
    CONSTRAINT dim_cliente_pkey PRIMARY KEY
(id_cliente)
)
WITH (
    OIDS = FALSE
)
TABLESPACE pg_default;

ALTER TABLE public.dim_cliente
    OWNER to postgres;
```



### **Tabela dim\_codigo\_postal**

```
CREATE TABLE public.dim_codigo_postal
(
    id_codigo_postal integer NOT NULL,
    nome_codigo_postal integer NOT NULL,
    CONSTRAINT dim_codigo_postal_pkey PRIMARY KEY
(id_codigo_postal)
)
WITH (
    OIDS = FALSE
)
TABLESPACE pg_default;

ALTER TABLE public.dim_codigo_postal
    OWNER to postgres;
```

### **Tabela dim\_date**

```
CREATE TABLE public.dim_date
(
    "DateKey" integer NOT NULL,
    "FullDate" date NOT NULL,
    "DayOfMonth" integer NOT NULL,
    "WeekDay" integer NOT NULL,
    "DayName" text COLLATE pg_catalog."default" NOT
NULL,
    "MonthName" text COLLATE pg_catalog."default" NOT
NULL,
    "Month" integer NOT NULL,
    "Year" integer NOT NULL,
    "WeekOfYear" integer NOT NULL,
    CONSTRAINT dim_date_pkey PRIMARY KEY ("DateKey")
)
WITH (
    OIDS = FALSE
)
TABLESPACE pg_default;

ALTER TABLE public.dim_date
    OWNER to postgres;
```

### **Tabela dim\_idade**

```
CREATE TABLE public.dim_idade
(
    id_idade integer NOT NULL,
    nome_idade integer,
    CONSTRAINT dim_idade_pkey PRIMARY KEY (id_idade)
)
WITH (
    OIDS = FALSE
)
TABLESPACE pg_default;

ALTER TABLE public.dim_idade
    OWNER to postgres;
```

### **Tabela dim\_status**

```
CREATE TABLE public.dim_status
(
    id_status integer NOT NULL,
    nome_status text COLLATE pg_catalog."default" NOT
NULL,
    CONSTRAINT dim_status_pkey PRIMARY KEY (id_status)
)
WITH (
    OIDS = FALSE
)
TABLESPACE pg_default;

ALTER TABLE public.dim_status
    OWNER to postgres;
```



## **ANEXO 5**

### **Código SQL das queries utilizadas no PRD**

## **Código SQL das queries utilizadas no Pentaho Report Designer**

### **Query principal para obtenção dos dados do relatório**

```
Select
    dim_centro.nome_centro,
    dim_status.nome_status,
    dim_cliente.nome_cliente,
    Avg(dim_idade.nome_idade) As "Tempo Médio Vida",
    Max(dim_idade.nome_idade) As "Tempo Max Vida",
    dim_date."MonthName",
    Count(fact_cartas.id_carta) As "Total Cartas"
From
    fact_cartas Inner Join
    dim_centro
        On fact_cartas.centro = dim_centro.id_centro Inner
Join
    dim_cliente
        On fact_cartas.cliente = dim_cliente.id_cliente
Inner Join
    dim_codigo_postal
        On fact_cartas.codigo_postal =
dim_codigo_postal.id_codigo_postal Inner Join
    dim_date
        On fact_cartas.data_imp_datekey =
dim_date."DateKey" Inner Join
    dim_idade
        On fact_cartas.idade = dim_idade.id_idade Inner
Join
    dim_status
        On fact_cartas.status = dim_status.id_status
WHERE
    dim_centro.nome_centro = ${enter_centro}
    AND dim_date."MonthName" = ${enter_mes}
GROUP BY
```

```
dim_centro.nome_centro,  
dim_status.nome_status,  
dim_cliente.nome_cliente,  
dim_date."MonthName"
```

### **Query utilizada para a passagem de parâmetros da variável Centro**

```
Select  
    dim_centro.nome_centro  
From  
    dim_centro  
Group By  
    dim_centro.nome_centro
```

### **Query utilizada para a passagem de parâmetros da variável Mês**

```
Select  
    dim_date."MonthName"  
From  
    dim_date  
Where  
    (dim_date."Month" = 4) Or  
    (dim_date."Month" = 5)  
Group By  
    dim_date."MonthName", dim_date."Month"
```



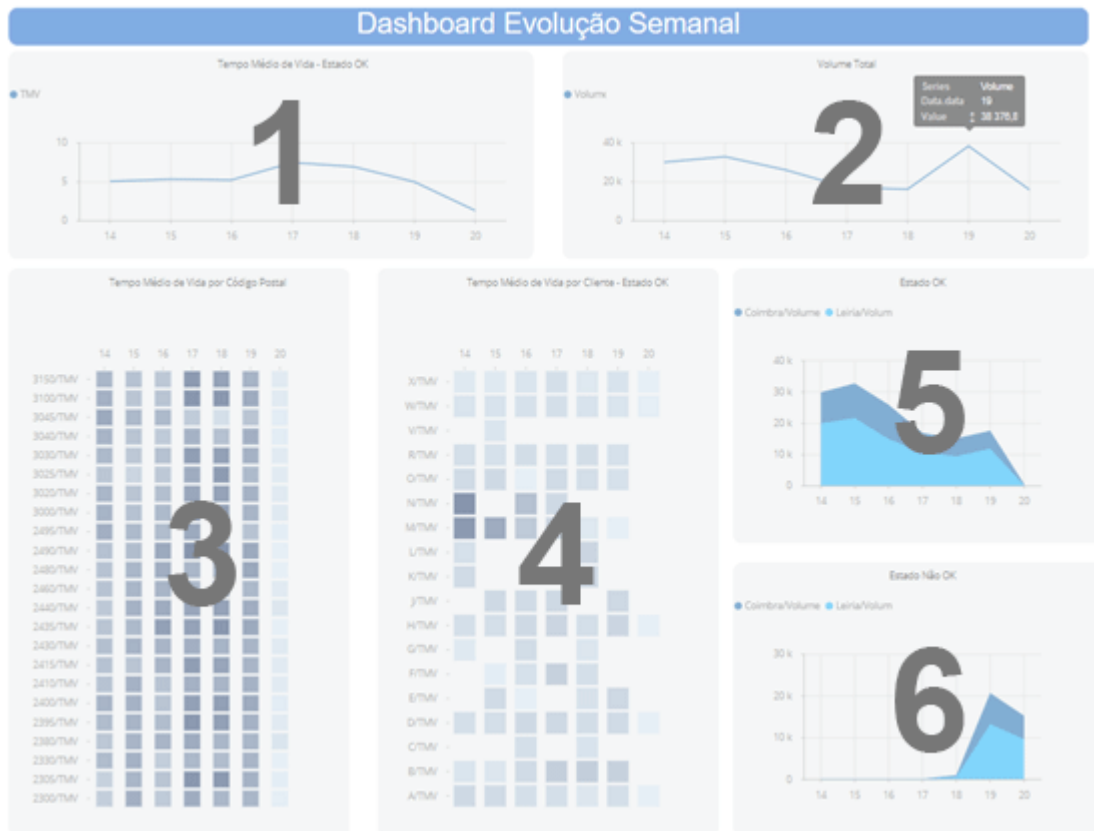
## **ANEXO 6**

### **Código MDX das queries utilizadas no CDE**



## Queries MDX Dashboard

Cada painel tem associada uma querye MDX sobre o cubo OLAP presente na Pentaho. De seguida são apresentadas a queries de cada painel de acordo com a figura abaixo.



### Query MDX painel 1

WITH

SET [~FILTER] AS

{[estado].[OK]}

SET [~ROWS] AS

{[data].[Semana].Members}

SELECT

NON EMPTY {[Measures].[TMV]} ON COLUMNS,

NON EMPTY [~ROWS] ON ROWS

```
FROM [Good Mailman]
WHERE [~FILTER]
```

### **Query MDX painel 2**

```
WITH
SET [~FILTER] AS
    {[estado].[estado].Members}
SET [~ROWS] AS
    {[data].[Semana].Members}
SELECT
NON EMPTY {[Measures].[Volume]} ON COLUMNS,
NON EMPTY [~ROWS] ON ROWS
FROM [Good Mailman]
WHERE [~FILTER]
```

### **Query MDX painel 3**

```
WITH
SET [~COLUMNS] AS
    {[codigo postal].[Código Postal].Members}
SET [~ROWS] AS
    {[data].[Semana].Members}
SELECT
NON EMPTY CrossJoin([~COLUMNS], {[Measures].[TMV]}) ON
COLUMNS,
NON EMPTY [~ROWS] ON ROWS
```

```
FROM [Good Mailman]
```

#### **Query MDX painel 4**

```
WITH
```

```
SET [~FILTER] AS
```

```
{[estado].[OK]}
```

```
SET [~COLUMNS] AS
```

```
{[cliente].[Cliente].Members}
```

```
SET [~ROWS] AS
```

```
{[data].[Semana].Members}
```

```
SELECT
```

```
NON EMPTY CrossJoin([~COLUMNS], {[Measures].[TMV]}) ON  
COLUMNS,
```

```
NON EMPTY [~ROWS] ON ROWS
```

```
FROM [Good Mailman]
```

```
WHERE [~FILTER]
```

#### **Query MDX painel 5**

```
WITH
```

```
SET [~FILTER] AS
```

```
{[estado].[OK]}
```

```
SET [~COLUMNS] AS
```

```
{[centro].[Centro].Members}
```

```
SET [~ROWS] AS
```

```
{[data].[Semana].Members}
```

```
SELECT  
  
NON EMPTY CrossJoin([~COLUMNS], {[Measures].[Volume]}) ON  
COLUMNS,  
  
NON EMPTY [~ROWS] ON ROWS  
  
FROM [Good Mailman]  
  
WHERE [~FILTER]
```

### **Query MDX painel 6**

```
WITH  
  
SET [~FILTER] AS  
  
    {[estado].[Not OK]}  
  
SET [~COLUMNS] AS  
  
    {[centro].[Centro].Members}  
  
SET [~ROWS] AS  
  
    {[data].[Semana].Members}  
  
SELECT  
  
NON EMPTY CrossJoin([~COLUMNS], {[Measures].[Volume]}) ON  
COLUMNS,  
  
NON EMPTY [~ROWS] ON ROWS  
  
FROM [Good Mailman]  
  
WHERE [~FILTER]
```