

**CENTRO PAULA SOUZA**

GOVERNO DO ESTADO DE  
**SÃO PAULO**

**Faculdade de Tecnologia de Americana  
Curso Superior de Tecnologia em Análise e Desenvolvimento de  
Sistemas**

# **Business Intelligence: visualização de dados utilizando a ferramenta open-source Pentaho**

**Isabela Muniz**

**Americana, SP  
2013**

# **Business Intelligence: visualização de dados utilizando a ferramenta open-source Pentaho**

**Isabela Muniz**  
Isabela.muniz@fatec.sp.gov.br

Trabalho Monográfico, desenvolvido em cumprimento à exigência curricular do Curso de Bacharelado em Análise de Sistemas e Tecnologia da Informação da Fatec-Americana, sob orientação do Prof. José Alberto Rodrigues.

**Área: Banco de dados,**

**Americana, SP  
2013**

# **Business Intelligence: visualização de dados utilizando a ferramenta open-source Pentaho**

Monografia aprovada como requisito parcial para obtenção do título de Bacharel em Análise de Sistemas e Tecnologia da Informação no curso de Análise de Sistemas e Tecnologia da Informação da Faculdade de Tecnologia de Americana.

## **BANCA EXAMINADORA**

**Professor Orientador:** Prof. José Alberto Rodrigues

**Professor:**

**Professor:**

## **AGRADECIMENTOS**

Antes de tudo agradeço a Deus, por deixar sempre a luz acesa para que eu pudesse encontrar o caminho certo.

Agradeço à minha mãe Antonia Cecília de Lima Muniz, por ser meu grande espelho e por ter feito de mim alguém que sabe sonhar e realizar sem perder a fé e esperança. Ao meu padrasto Ataíde por ser meu exemplo de trabalhador dedicado e incansável.

Ao meu irmão, Vinicius de Lima Muniz, por ter me incentivado a entrar na área de TI e por ser meu suporte durante esses quatro anos de formação, tanto com ajudas técnicas quanto com momentos divertidos e agradáveis dos quais sempre me lembro.

Ao meu orientador, profº Beto, que aceitou a minha proposta de monografia e me auxiliou a concluir esse trabalho.

A meus amigos de sala que estiveram comigo em estudos madrugadas adentro e também em momentos de descontração, me proporcionando um ambiente onde pude descobrir muito sobre o convívio em sociedade e, mais ainda, sobre mim.

A meus ex-colegas de trabalho da CATI, onde realizei meu primeiro estágio, aprendi muito e fiz amigos que nunca me esquecerei, e a meus colegas de trabalho da Ícaro Technologies, por me ensinarem muito e participarem de forma especial da minha vida.

## DEDICATÓRIA

Dedico este trabalho a Deus, aos meus pais, pelo apoio e paciência, aos meus amigos que aqui fiz e que eternamente estarão em minhas lembranças.

## RESUMO

Com o mercado cada vez mais competitivo torna-se essencial que as empresas de todos os setores busquem organizar e analisar seus dados de forma criteriosa, a fim de detectar quais os pontos críticos do negócio e, então, tornar mais eficaz a tomada de decisão por parte dos gestores. Nesse contexto, Business Intelligence aparece como uma solução eficiente, com a estruturação da base de dados e posterior prospecção de informações importantes a quem realmente necessita delas, mas também pode demandar grandes investimentos se a opção for por utilizar uma ferramenta proprietária. Esse trabalho apresenta a ferramenta Pentaho Report Designer, que oferece diversos recursos na criação de relatórios e é livre (Open Source), minimizando os custos de implementação de BI em uma organização.

**Palavras Chave:** Business Intelligence. Visualização de dados. Open Source.

## **ABSTRACT**

In the current increasingly competitive market it is essential that companies from all sectors seek to organize and analyze their data to detect critical points of the business and then make more effective decision-making management. In this context, Business Intelligence appears as an efficient solution, with the structuring of the database and subsequent exploration of important information to those who really need them, but can also require large investments if the option is to use a proprietary tool. This paper will present the Pentaho Report Designer tool, which offers many features to create reports and is free (Open Source), minimizing the costs of implementing BI in an organization.

**Keywords:** Business Intelligence. Data visualization. Open Source.

## SUMÁRIO

<b>1. INTRODUÇÃO.....</b>	<b>11</b>
<b>2. BUSINESS INTELLIGENCE .....</b>	<b>13</b>
2.1. OLTP .....	17
2.2. ETL (Extract, Transform, Load) .....	17
2.3. Data Warehouse.....	18
2.4. OLAP .....	19
2.5. Visualização de dados.....	20
<b>3. PENTAHO .....</b>	<b>23</b>
3.1. Pentaho Report Designer (PRD) .....	25
3.1.1 Fontes de dados .....	25
3.1.2 Recursos gráficos.....	26
3.1.3 Agregações e campos calculados.....	27
3.1.4 Solicitação de parâmetros.....	27
3.1.5 Relatórios dinâmicos e Formatação Condicional .....	27
3.2. Pentaho BI Server .....	27
<b>4. UTILIZANDO O PRD.....</b>	<b>29</b>
4.1. Criação do Relatório.....	31
4.2. Publicando o relatório no BI Server .....	36
4.3. Resultado .....	38
<b>5. CONSIDERAÇÕES FINAIS.....</b>	<b>40</b>
<b>6. REFERÊNCIAS .....</b>	<b>42</b>



## Lista de Figuras

Figura 1 - Etapas de Implementação de um sistema de BI.....	14
Figura 2 - Ambiente de Business Intelligence .....	16
Figura 3 - Relatório, Dashboard e Cubo OLAP .....	21
Figura 4 - Borboleta de BI com Pentaho .....	24
Figura 5 - Diagrama do modelo dimensional de dados da Beltrano S/A.....	30
Figura 6 - Interface inicial do PRD.....	31
Figura 7 - Cabeçalho do Relatório de Treinamentos.....	32
Figura 8 - Criação de tabela do Relatório de Treinamentos.....	33
Figura 9 - Função de agregação no Relatório de Treinamentos .....	35
Figura 10 - Criação de gráfico no Relatório de Treinamentos.....	35
Figura 11 - Relatório de Treinamentos.....	36
Figura 12 - Conexão do PRD com o BI Server.....	37
Figura 13 - Publicação do relatório no BI Server.....	38

**Lista de Tabelas**

Tabela 1 - Arquiteturas de OLAP .....20

## 1. INTRODUÇÃO

Desde o marco da Revolução Industrial surgiram diversos avanços tecnológicos, gerando um aumento de produtividade e, por consequência, um aumento de capital. Com isso, tornou-se essencial analisar as informações e suas mudanças, para orientar recursos de forma que fosse possível produzir de forma mais eficiente, em menos tempo possível e com qualidade.

A tecnologia foi avançando por esse caminho chegando aos anos 1990, quando grandes organizações utilizavam apenas Centros de Informações (CI) e Centro de Processamento de Dados (CPD). A partir daí foi detectada a necessidade de avançar quanto ao armazenamento de dados e, por volta de 1992/1993, surgiu o Data Warehouse.

Já há algum tempo as corporações têm preocupações além de otimizar seus processos de produção. Com o passar dos tempos, os hábitos, comportamentos e consumo das pessoas mudou muito, e o ambiente empresarial, como não poderia deixar de ser, foi fortemente afetado por essas mudanças. As empresas passaram a se preocupar ainda mais com a análise das informações disponíveis, pois esse contexto deixou clara a necessidade de manterem-se proativas, prevendo o comportamento futuro de seus clientes e do mercado.

A concorrência entre as organizações é cada vez mais acirrada, e para manter uma posição entre as que sobrevivem à pressão, é preciso investir em vários aspectos, dentre eles, necessariamente, a tecnologia da informação. É nesse contexto que as ferramentas de Business Intelligence aparecem como uma forma eficaz de auxílio à tomada de decisões para a continuidade da organização em seu meio.

O capítulo 2 apresenta o referencial teórico necessário para o entendimento do conceito de Business Intelligence, incluindo toda a estrutura de uma solução que se utilize do mesmo.

Após isso, o capítulo 3 realiza uma introdução ao Pentaho enquanto empresa, e então às ferramentas Pentaho Reporting Design e Pentaho BI Server e seus recursos disponível. Dessa forma, no capítulo 4 é realizada uma experimentação dessas ferramentas no processo de criação e disponibilização de um relatório que

deve ser útil na definição de uma estratégia para aumento de receita de uma empresa fictícia, chamada Beltrano S/A.

Por fim, o capítulo 5 apresenta as considerações finais a cerca de todo o desenvolvimento do trabalho; enquanto o capítulo final lista as referências utilizadas na confecção deste.

## 2. BUSINESS INTELLIGENCE

Segundo Turban et. al. (2009), é possível definir BI como um termo "guarda-chuva" que contempla arquiteturas, metodologias, bancos de dados, ferramentas, e aplicações buscando permitir o acesso interativo aos dados, proporcionar a manipulação dos mesmos e fornecer aos gerentes e analistas de negócios a capacidade de realizar a análise adequada para a tomada de decisão. Como tomada de decisão entende-se muito mais do que o momento da escolha da ação, é um processo que envolve processos comparativos, escolha de parceiros, exame da capacidade operacional da empresa, análise de cenários, projeção de futuro, identificação de riscos e de oportunidades, análise da concorrência, dentre outros tantos requisitos (FRANCO, 2009).

Considerando esses conceitos é possível dizer que sistemas de BI possibilitam que as organizações tenham uma visão abrangente sobre os pontos que impactam seu negócio, com a definição e apresentação de métricas com uma versão absoluta (sem margens para diferentes interpretações), e em tempo hábil. A capacidade de fornecer informações relevantes quando for preciso (inclusive em tempo real) é que torna o BI um diferencial.

O processo de BI baseia-se na transformação de dados em informações. (TURBAN et. al, 2009). O dado não tem significado relevante a princípio e, sendo assim, não tem valor para justificar ou respaldar conclusões. A informação é a organização dos dados de forma a agregar significado e entendimento dentro de um contexto determinado.

BI engloba processos, pessoas, culturas, gestão da informação, negócios e muitos outros aspectos. Qualquer área de uma empresa pode se beneficiar do uso de dados para agilizar processos, melhor atender as necessidades dos clientes e usuários, tornando-se mais eficaz e mais competitiva. Apesar disso, ainda há alguma resistência na utilização de BI por parte de pequenas e médias empresas que acreditam não possuem capital suficiente para sustentar financeiramente uma solução. Entretanto, o investimento tem relação com o crescimento da empresa, podendo ser moderado e acessível de início.

Para que a implementação da solução de BI tenha efetividade é preciso que seja entendida cada uma de suas etapas. Além disso, é necessário que a solução

seja eficiente, fornecendo informações inteligentes para os tomadores de decisão, e eficaz, atingindo o objetivo de forma precisa. (ELIAS (2013)).

O diagrama disponível na Figura 1 traz uma visão macro das atividades que compõe o processo de implementação de um sistema de BI.

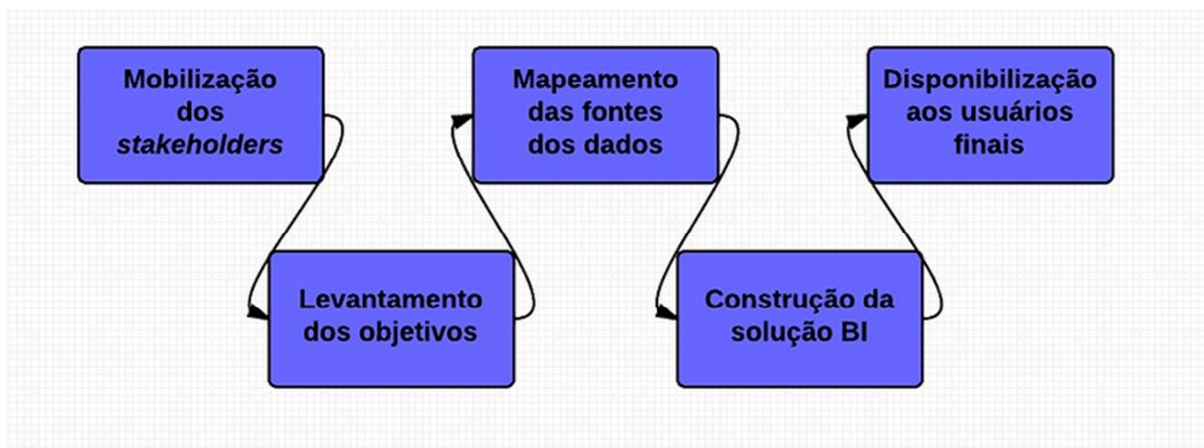


Figura 1 - Etapas de Implementação de um sistema de BI  
Fonte: BI na Prática<sup>1</sup>

A primeira etapa, definida como Mobilização dos *stakeholders*, consiste em identificar as partes interessadas do processo de BI. É necessário reconhecer quem são os beneficiados da solução, além de buscar o apoio total da alta gestão. Os benefícios que o projeto trará para a organização devem ser apresentados, além de promover discussões sobre os diversos assuntos relacionados à gestão através de indicadores.

A segunda atividade é definida como Levantamento dos Objetivos, quando serão definidas as necessidades de informação que o gestor possui. Nessa etapa é importante buscar respostas para a questão “Quais informações necessito para tomar decisões mais efetivas?”, ainda sem preocupação a respeito das fontes de dados.

Após isso, a terceira etapa será a busca pelos dados já existentes na organização, oriundos de diversas fontes (bancos de dados, planilhas, mídias físicas entre outras coisas). Com a descoberta de dados importantes, deve ser feito um planejamento para acesso aos mesmos; e se for detectado que dados necessários não estão disponíveis, é preciso definir uma estratégia para descobri-los.

<sup>1</sup> Disponível em: <<http://www.binapratca.com.br/#!etapas-bi/c1akd>>. Acesso em: 20 out. 2013.

A quarta etapa é, normalmente, a maior dentre todas. A construção da solução de BI é o momento onde é feita a modelagem, o processo de ETL (extração, transformação e carga) e então a base consolidada do Data Warehouse é populada.

Por fim, é necessário disponibilizar a solução. É importante que os usuários dessa solução sejam treinados para extrair o melhor dela.

Segundo Silva (2013), considerando todo o investimento, é possível destacar benefícios significativos de uma solução de BI em uma organização:

“Redução de processos.

Substitui outras ferramentas, como por exemplo, Excel, no sentido de desenvolver fórmulas no mesmo para busca de informações, bem como de outros sistemas mais “engessados” em termos de gerar informações e relatórios.

Agilidade na obtenção de informações e relatórios.

Flexibilidade e personalização de informações e relatórios, facilitando as análises a serem realizadas.

Facilidade de modificação/melhoria nos relatórios, como por exemplo, alterar a forma de apresentar os dados e acrescentar novas informações contribuindo assim para a evolução contínua das análises de informações como também da tecnologia.

Disponibilização de vários relatórios para confrontar os dados entre os diversos módulos do próprio B.I.

As informações são disponibilizadas na tela do software em poucos segundos dependendo dos filtros (parâmetros) realizados.

Facilita o planejamento da empresa nos diversos níveis, não só no nível estratégico, como também no tático e operacional.

Acesso às informações praticamente on-line.”<sup>2</sup>

Com relação à arquitetura, um sistema de BI pode ser implementado de diversas maneiras, porém a estrutura mais comum é a composta por uma ferramenta de ETL, uma tecnologia de Data Warehouse e ferramentas OLAP (*Online Analytical Processing*). A Figura 2 ilustra essa arquitetura.

---

<sup>2</sup> Disponível em: <<http://www.dm.com.br/texto/125545-business-intelligence-a-inovacao-para-tomada-de-decisoes-organizacionais>>. Acesso em: 20 ago. 2013

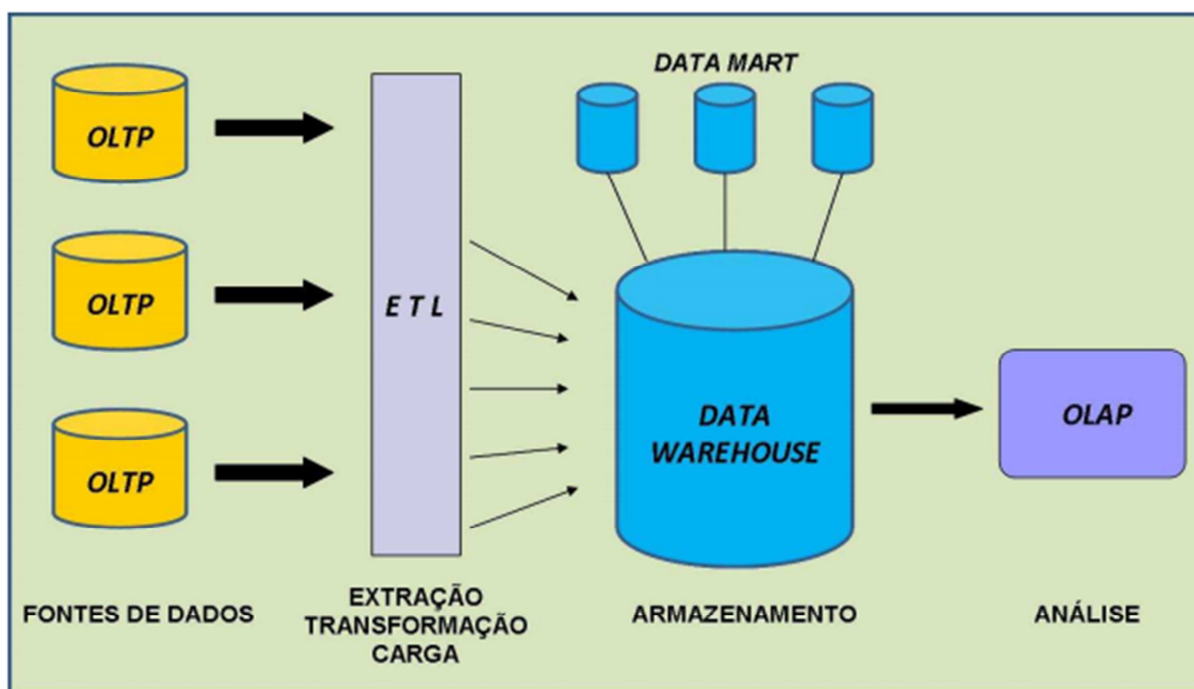


Figura 2 - Ambiente de Business Intelligence  
 Fonte: Adaptação segundo (BARBIERI, 2001).

As fontes de dados, ou OLTPs representam os bancos de dados transacionais, de onde são extraídos os dados a serem tratados através da ferramenta de ETL (Extract, Transform, Load, em português Extração, Transformação, Carga de dados). Após isso, esses dados devem ser modelados de acordo com as definições iniciais, e então carregados no Data Warehouse, onde ocorrerá o armazenamento das transações e atualizações realizadas nas fontes de dados. Os Data Mart's apresentados na Figura 2 são como data warehouses específicos para seções da empresa. Por fim, as ferramentas de OLAP são empregadas para realizar a análise dos dados armazenados no Data Warehouse.

Após apresentar a relação de cada componente na estrutura final, os próximos tópicos irão apresentar mais detalhes a respeito de cada um, assim como a importância dos mesmos individualmente. Após isso, será definido o conceito de visualização de dados e sua relevância dentro de uma solução de Business Intelligence.



## 2.1. OLTP

OLTP (Online Transaction Processing) é Processamento de Transações em Tempo-Real e pode ser definido como um sistema encarregado de registrar todas as transações que fazem parte de uma operação organizacional. São os bancos de dados tradicionais, incapazes de gerar relatórios para uma análise de dados da empresa. Por essa razão, torna-se necessária a criação do Data Warehouse (KIMBALL, 1998).

Os sistemas OLTP permitem a inserção, consulta, atualização e deleção de dados e são norteados pelos processos desenvolvidos na empresa, como Pedido, Pagamento, Nota e - diferente do Data Warehouse - são voláteis, passando por constantes modificações dos dados (INMON, 1997).

A consistência do OLTP é microscópica, tendo a finalidade de processar grandes quantidades de transações minúsculas e atômicas, sem perder nenhuma delas. A essência de uma transação é que tanto o remetente quanto o destinatário estão sempre de acordo se esta foi concretizada (KIMBALL, 1998).

Esses bancos são a fonte de extração de dados do data warehouse.

## 2.2. ETL (Extract, Transform, Load)

A padronização de dados é fundamental para que o Data Warehouse possa processá-los. Os dados provenientes das várias OLTPs não são uniformemente formatados, e então se torna necessário utilizar uma ferramenta ETL, que como o próprio nome diz, realiza a extração, tratamento e carga de dados.

Extração: é a coleta de dados dos sistemas de origem (fontes de dados), extraíndo-os e transferindo-os para o ambiente de Data Warehouse.

Transformação: nesta etapa são aplicadas diversas regras e/ou funções para realizar os ajustes necessários visando melhorar a qualidade dos dados e consolidá-los independentemente de suas origens. O volume de transformações as quais os dados serão submetidos tem muita ligação com a fonte dos mesmos. O objetivo é padronizar.

Carga dos dados: consiste em estruturar e carregar os dados fisicamente para a camada de apresentação seguindo o modelo dimensional. A latência e o alcance de reposição ou acréscimo dependerão das estratégias traçadas no projeto, dependendo de tempo disponível e necessidades do negócio.

Segundo Inmon (2002) apud CRAMER (2006, p. 21): "O processo de integração e transformação de dados, tipicamente, consome até 80% dos recursos de desenvolvimento".

### **2.3. Data Warehouse**

A tecnologia da informação possui o objetivo de transformar dados históricos em conhecimento estratégico nas organizações. O DW (Data Warehouse) é uma ferramenta que surgiu com esse propósito, trazendo o conceito de centralização das informações, visualização multidimensional dos dados, além da descoberta de padrões de comportamento, dando aos administradores maior agilidade na tomada de decisões (BARBALHO, 2003).

De forma direta, Kimball (1998) define o Data Warehouse como "uma cópia de dados de transação (dados transacionais dos sistemas de origem), especificamente estruturado para consulta e análise".

O DW é o alicerce de sistemas de BI, tornando a tarefa de um analista de BI imensamente menor do que em um ambiente clássico. O DW é um conjunto de dados integrado (dados a partir de uma variedade de origens e fundidos em um todo coerente), não volátil (dados são adicionados, mas nunca excluídos), variável em relação ao tempo (todos os dados no armazém de dados são identificados com um período de tempo particular) e baseado em assuntos (os dados que dão informações sobre um assunto particular em vez de sobre operações contínuas da companhia), servindo de apoio à decisões gerenciais (INMON, 1997).

O DW sempre irá armazenar o elemento tempo (ano, mês, semana, dia...), para possuir um registro do momento em que uma transação foi realizada. Geralmente é utilizado um histórico de informações entre cinco a dez anos. (INMON, 1997).

A questão da granularidade no Data Warehouse não é algo pré-determinado por padrão, mas deve ser definida no início do projeto. É possível construir um DW com granularidade alta (baixo nível de detalhe) ou baixa (alto nível de detalhe). As duas opções apresentam vantagens de acordo com as necessidades de informação e infraestrutura do ambiente. Com granularidade alta são consumidos menos recursos computacionais, e as consultas são otimizadas em tempo. Por outro lado, uma granularidade baixa consegue responder perguntas muito específicas, pois irá armazenar todos os registros possíveis.

Contudo, é bastante incomum que seja examinado um evento isolado em um ambiente de DW; é mais comum que ocorra a utilização de uma visão de conjunto dos dados, o que significa que um grande volume de registros será examinado. Nesse contexto, tanto a granularidade baixa quanto a alta conseguem satisfazer à necessidade, mas a utilização de recursos na baixa será gigante, enquanto num nível alto os dados são mais compactos. O uso dos dados do nível alto de granularidade é muito mais eficiente se ele apresenta o detalhamento suficiente (INMON, 1997).

## **2.4. OLAP**

Criado em 1993, por Edgar F. Codd, o OLAP (Online Analytical Processing, em português Processamento Analítico em Tempo Real) é uma ferramenta que permite aos analistas de negócios, gerentes e executivos realizar a análise e visualização de dados corporativos em tempo hábil, de forma consistente e principalmente interativa.

Em 1995 foi criado o The OLAP Council, um conselho que padroniza a tecnologia OLAP, e define o termo como uma categoria de tecnologia de software que permite aos analistas, gerentes e executivos obter entendimento sobre os dados de forma rápida e consistente, através de um acesso interativo a uma grande variedade de possíveis pontos de vista de informações que foram transformadas a partir de dados brutos que refletem a real dimensionalidade da empresa no entendimento do usuário.

Existem quatro arquiteturas diferentes de OLAP, classificadas em ROLAP (Relational On Line Analytical Processing), MOLAP (Multidimensional On Line Analytical Processing), HOLAP (Hybrid On Line Analytical Processing) e DOLAP (Desktop On Line Analytical Processing).

Tabela 1 - Arquiteturas de OLAP

Arquitetura	Características
<b>ROLAP</b>	<ul style="list-style-type: none"> <li>Os dados são armazenados em banco de dados relacionais;</li> <li>A manipulação dos dados armazenados no banco de dados relacional é feita para dar a aparência de operação Slice/Dice tradicional;</li> <li>Na essência, cada ação de Slice/Dice é equivalente a adicionar uma cláusula WHERE em uma declaração SQL.</li> </ul>
<b>MOLAP</b>	<ul style="list-style-type: none"> <li>Arquitetura OLAP tradicional;</li> <li>Os dados são armazenados em cubos dimensionais, em formatos proprietários, e não em banco de dados relacionais;</li> <li>O usuário trabalha, monta e manipula os dados do cubo diretamente no servidor.</li> </ul>
<b>HOLAP</b>	<ul style="list-style-type: none"> <li>Tenta combinar as vantagens de MOLAP e ROLAP, extraindo o que há de melhor de cada uma, ou seja, a alta performance do MOLAP com a melhor escalabilidade do ROLAP;</li> <li>Para informações do tipo síntese, HOLAP utiliza cubos dimensionais para um desempenho mais rápido;</li> <li>Quando for necessário mais detalhe de uma informação, pode ir além do cubo multidimensional para o banco de dados relacional utilizado no armazenamento dos detalhes.</li> </ul>
<b>DOLAP</b>	<ul style="list-style-type: none"> <li>A partir de um cliente qualquer emitem uma consulta para o servidor e recebem o cubo de informações de volta para ser analisado na estação cliente.</li> </ul>

Fonte: Rogério Araújo<sup>3</sup>

## 2.5. Visualização de dados

A visualização de dados foi definida pela empresa Atom BI (s.d.) como:

<sup>3</sup> Disponível em: <<http://rogerioaraujo.wordpress.com/2009/05/27/bi-arquiteturas-olap/>>. Acesso em: 25 ago. 2013

“uma técnica utilizada para comunicar a informação de maneira clara e ordenada utilizando meios gráficos, e abrange um conjunto de melhores práticas que incluem atributos como tipo de gráficos, tabelas, cores, tamanho, tipo de fontes e posicionamento de elementos visuais que faz com que as pessoas vejam, utilizem e interajam com as informações de uma forma simples e eficiente.[...]”<sup>4</sup>

Considerando que o principal objetivo do BI seja fornecer informações relevantes que tenham impacto nas tomadas de decisão gerenciais, a visualização de dados se torna um elemento chave, uma vez que, para os gestores, será este o elemento utilizado de fato; todas as etapas anteriores, que envolvem tratar os dados e relacioná-los, culminam nesse ponto.

Quando se fala em visualização dados, as apresentações mais comuns são em forma de relatórios, dashboards e cubos OLAP. A Figura 3 ilustra um exemplo de cada um deles.

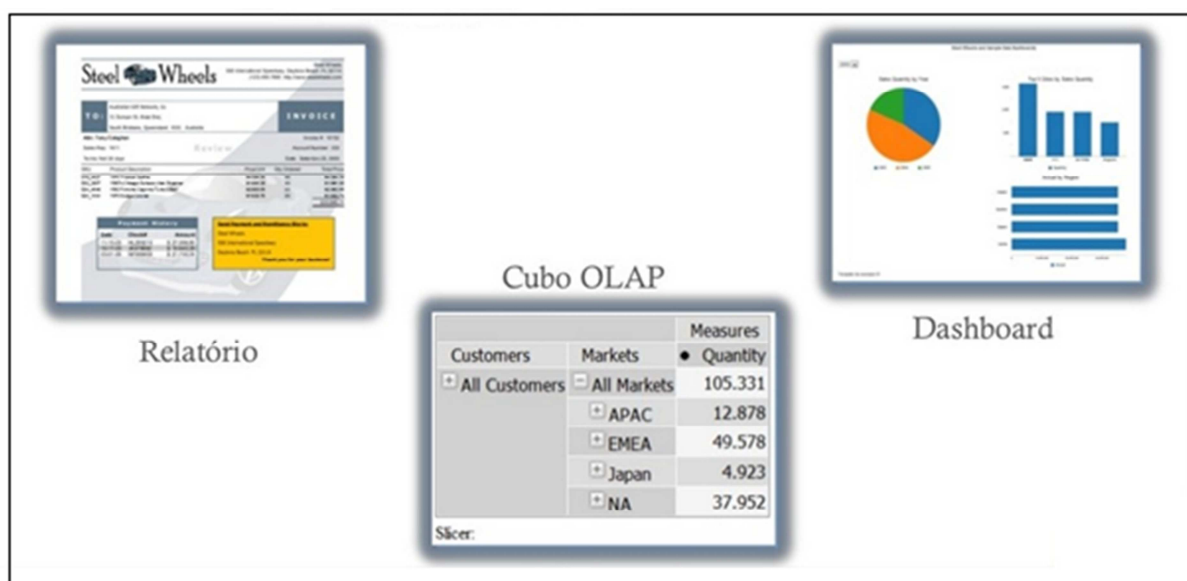


Figura 3 - Relatório, Dashboard e Cubo OLAP  
Fonte: IT4biz<sup>5</sup>

Relatórios podem envolver dados relevantes de toda a organização, ou de áreas específicas, utilizando tabelas, gráficos, ou mesmo informações organizadas de forma textual.

Em cubos OLAP as informações são organizadas em duas ou mais

<sup>4</sup> Disponível em: <<http://www.atomsail.com/pt/o-que-e-visualizacao-de-dados-saas-atomsail.php>>. Acesso em: 01 out. 2013.

<sup>5</sup> Disponível em: <<http://www.slideshare.net/caiomsouza/curso-gratuito-online-desmistificando-bi-business-intelligence-open-source-pentaho-e-tecnologias-livres>>. Acesso em: 01 out. 2013.

dimensões, onde cada dimensão pode ser qualquer visão de negócio significativa para o negócio. Dentro de cada dimensão desse modelo os dados são organizados hierarquicamente, possibilitando diversos níveis de detalhes.

Dashboards são apresentações visuais das informações mais importantes e necessárias para alcançar um ou mais objetivos de negócio, consolidadas e ajustadas em uma única tela para fácil acompanhamento do seu negócio.

Para construir uma apresentação de dados clara e objetiva é necessária a utilização de uma ferramenta especializada, com recursos que facilitem o agrupamento de informações e, assim, tornem seu entendimento mais fácil. Existem diversas ferramentas disponíveis no mercado para esse fim, tanto proprietárias quanto Open Source.

Uma das ferramentas Open Source que se destaca nessa área é a Pentaho Report Design, destinada à criação de relatórios. O próximo capítulo descreve quais são seus recursos disponíveis, que a fazem ganhar espaço entre outras ferramentas com a mesma finalidade.

### 3. PENTAHO

Pentaho é uma empresa fundada em 2004 por cinco profissionais de BI, que desenvolveram ferramentas que cobrem toda uma solução da área. Atualmente são disponibilizadas versões Enterprise (proprietárias, que requerem a compra de licenças) e Community (Open Source). Com o tempo, essas ferramentas passaram a ganhar espaço no mercado, sendo utilizadas por grandes empresas, e em 2013 seu nome já é encontrado no *Magic Quadrant* (quadrante mágico) de plataformas de BI do Gartner.

O quadrante do Gartner divide as empresas que se destacam no mercado de BI a partir da parametrização por habilidade de execução da plataforma *versus* plenitude de visão. Com isso, há uma divisão entre quatro categorias: líderes, competidores, visionários e destaques em um nicho. O Pentaho se encontra na última categoria.

A suite de BI oferecida na versão Community (Open Source) pela Pentaho é uma das mais populares atualmente. A Pentaho BI Platform (plataforma de BI Pentaho) possui as ferramentas e tecnologias para construir a infraestrutura e os processos que habilitam os analistas de negócio a obter insights oportunos, precisos, valiosos e práticos sobre seus negócios. A Figura 4 apresenta a estrutura conhecida como "borboleta de BI", onde são encaixadas duas soluções do portfólio Pentaho, em diferentes momentos.



Figura 4 - Borboleta de BI com Pentaho  
Fonte: Geek for BI<sup>6</sup>

É possível realizar uma comparação da Figura 4 com a Figura 2, que ilustra a estrutura de uma solução de BI. Sendo assim, fica claro que existem opções de ferramentas além da visualização de dados, preenchendo a etapa de ETL (com o Pentaho Data Integration) e, após o armazenamento de dados no Data Warehouse, existe toda a plataforma de BI do Pentaho para análise, manipulação, disponibilização e visualização de dados.

Para visualização de dados, especificamente criação de relatórios, há uma suite chamada Pentaho Reporting, que contempla o Pentaho Report Designer (PRD), Pentaho Reporting Engine (PRE) e Pentaho Reporting SDK. Com o PRD é possível criar relatórios visualmente, enquanto os outros componentes são utilizados como integração a sistemas já existentes.

Além das opções dessa suite, é possível criar relatórios utilizando o Pentaho BI Server, de forma mais simples do que com o PRD; e também cubos OLAP, acessando a parte de analysis.

Os próximos tópicos irão detalhar os recursos disponíveis no PRD e na criação de relatórios pelo próprio Pentaho BI Server, além da integração entre ambas, a fim de demonstrar como as duas ferramentas podem criar relatórios completos, gerando uma visualização de dados esclarecedora.

<sup>6</sup> Disponível em: <<http://geekbi.wordpress.com/2013/10/29/de-volta-ao-basico-pentaho-bi/>>. Acesso em: 01 nov. 2013.



### 3.1. Pentaho Report Designer (PRD)

A PRD é uma aplicação desktop que proporciona um ambiente de design visual para criação de relatórios com diversos recursos, e direcionado a usuários familiarizados com as fontes de dados a serem utilizadas. A ferramenta está atualmente na versão 3.9, e é possível utilizá-la nos sistemas operacionais Linux, Windows e Mac OS. Por ser um software livre, para utilizá-lo é necessário somente realizar o download através de um link disponível na página oficial do Pentaho Community.

Os subcapítulos a seguir apresentam de forma mais detalhada seus recursos disponíveis.

#### 3.1.1 Fontes de dados

Com relação às fontes de dados, há muitas opções suportadas pela ferramenta. As opções mais comuns são:

- Bancos de dados relacionais utilizando drivers JDBC e ODBC
- Arquivos XML via queries XPath queries
- OLAP data-sources: acesso nativo a ferramenta de analysis do Pentaho, chamada Mondrian, além de XML/A e outras bases com OLAP4J.
- Acesso a views definidas em outra ferramenta chamada Pentaho Metadata System.

A cobertura de uma diversidade grande de fontes de dados é um recurso muito importante da ferramenta, uma vez que torna possível a visualização de dados em uma estrutura diferente da recomendada para soluções de BI. Caso a empresa não possua a arquitetura ideal, pode utilizar suas bases já existentes (desde que estas estejam organizadas e com alguma padronização) para retirar informações relevantes. Um grande diferencial da ferramenta, se comparada a muitas outras

disponíveis no mercado com a mesma finalidade, é a possibilidade de utilizar múltiplas fontes na criação de relatórios.

Focando na utilização de bancos de dados relacionais, há mais de trinta opções para utilização no PRD, o que cobre uma fatia muito grande das opções utilizadas atualmente. É possível configurar novas fontes de dados através da interface da própria ferramenta, porém é necessário adicionar a biblioteca correspondente à base entre as utilizadas pelo software.

### 3.1.2 Recursos gráficos

Há muitos recursos gráficos disponíveis na ferramenta, entre elementos simples e elaborados, possibilitando uma grande personalização dos relatórios. Abaixo são listados esses elementos:

- Legendas, campos de mensagens, número e data.
- Imagens estáticas e dinâmicas (nos formatos PNG, JPEG, GIF, WMF, SVG).
- Códigos de barras nos seguintes padrões: Code128, Code128A, Code128B, Code128C, UCC128, EAN128, EAN13, Bookland, UPC-A, Padrão 2 de 5, Interleave 2 de 5, PDF417, Código 39, Codabar, PostNet.
- Sparklines: Bar-Spark, Line-Spark, Pie-Spark.
- Gráficos: barra, linha, área, torta, Multi-Pie, Bar-Line, Ring, Bubble-Chart, Dispersão Gráfico, Radar, Cachoeira.
- Subseções dentro do relatório.

Além dessas opções, ainda é disponibilizada uma API (Application Programming Interface, ou Interface de Programação de Aplicativos) simples para criação de novos tipos de elementos.

### **3.1.3 Agregações e campos calculados**

A ferramenta suporta todos os tipos de agregação importantes (soma, média, contagem, mínimo e máximo). Além disso, são suportadas as seguintes opções:

- Expressões Open-fórmula de cálculo estilo Excel.
- Expressões script para cálculos complexos.
- API extensível para adicionar seus próprios cálculos.

### **3.1.4 Solicitação de parâmetros**

A ferramenta suporta a adição de parâmetros a serem inseridos pelo usuário. Esses parâmetros podem ser entradas comuns ou opções pré-definidas.

### **3.1.5 Relatórios dinâmicos e Formatação Condicional**

Às vezes, os relatórios estaticamente definidos não são suficientes; usuários avançados necessitam de opções mais dinâmicas. A flexibilidade do mecanismo de relatórios Pentaho permite ajustar praticamente todas as propriedades da definição do relatório.

A formatação condicional permite que a estrutura do relatório sofra alterações em tempo de execução, adaptando as informações com o máximo de flexibilidade.

## **3.2. Pentaho BI Server**

A implementação mais popular da plataforma de BI criada pelo Pentaho é o Pentaho BI Server, responsável por toda a camada servidora, fornecendo toda a

arquitetura e infraestrutura necessária para a construção de soluções para Business Intelligence, com nove serviços básicos:

- Autenticação;
- Logs;
- Auditoria;
- Serviços web;
- Motores de regras de negócio;
- Mecanismo para integração com relatórios;
- Mecanismo para integração de análises;
- Mecanismo para integração com dashboards;
- Mecanismo para integração com mineração de dados.

Sua arquitetura é composta por diversos outros softwares livres, como Apache Tomcat e Jetty como servlets, Spring Security para controle de autenticação e Hibernate para realizar o mapeamento das bases de dados para o datawarehouse.

Utilizando o Pentaho BI Server é possível criar relatórios relacionais e analíticos, com dados provenientes de diversas fontes e com a possibilidade de extração nos formatos PDF, Excel, HTML E CSV. A grande diferença entre a versão Enterprise (que é paga) e a versão Community (gratuita, utilizada nesse trabalho), além do preço, é que a versão Enterprise possui uma suíte de ferramentas já previamente integradas e algumas mais desenvolvidas que a versão livre. Além disso, contém suporte técnico, manutenção com atualizações frequentes, e acesso a melhor documentação e treinamento.

#### 4. UTILIZANDO O PRD

Após a apresentação dos recursos principais da ferramenta PRD, esse capítulo apresenta uma simulação de uso da mesma para criação de um relatório, e após isso, a publicação deste no Pentaho BI Server.

Para possibilitar essa experimentação é utilizada uma base de dados disponibilizada por Fábio de Salles, no site oficial do livro Pentaho na Prática. Essa base se refere à empresa Beltrano S/A, que é definida por Salles como:

*“[...] uma empresa fictícia, que desenvolve e vende treinamentos. Ela tem um catálogo de cursos, que são vendidos em turmas. Assim, quando um cliente compra algo, compra vagas em uma turma (e não vagas em um curso) que é de um determinado curso. Se um curso não tem uma turma aberta, então ele (o curso) não pode ser vendido. Cada vaga (ou lote de vagas) é comprada por um cliente que pode ser PF ou PJ, e é vendida por um empregado da Beltrano. Cada vaga é comercializada por um preço do catálogo, e pode ter ou não um desconto. Cada curso tem um autor, registrado na tabela de empregados.”<sup>7</sup>*

O gestor da Beltrano S/A gostaria de obter informações sobre os treinamentos, para pensar sobre uma estratégia para aumentar a receita da empresa. Para criar um relatório para esse fim, é necessário apresentar a base de dados da Beltrano S/A. Essa apresentação é realizada através da Figura 5.

---

<sup>7</sup> Disponível em: <<http://geekbi.wordpress.com/2013/09/27/base-de-treinamento-beltrano-sa/>>. Acesso em: 01 nov. 2013.

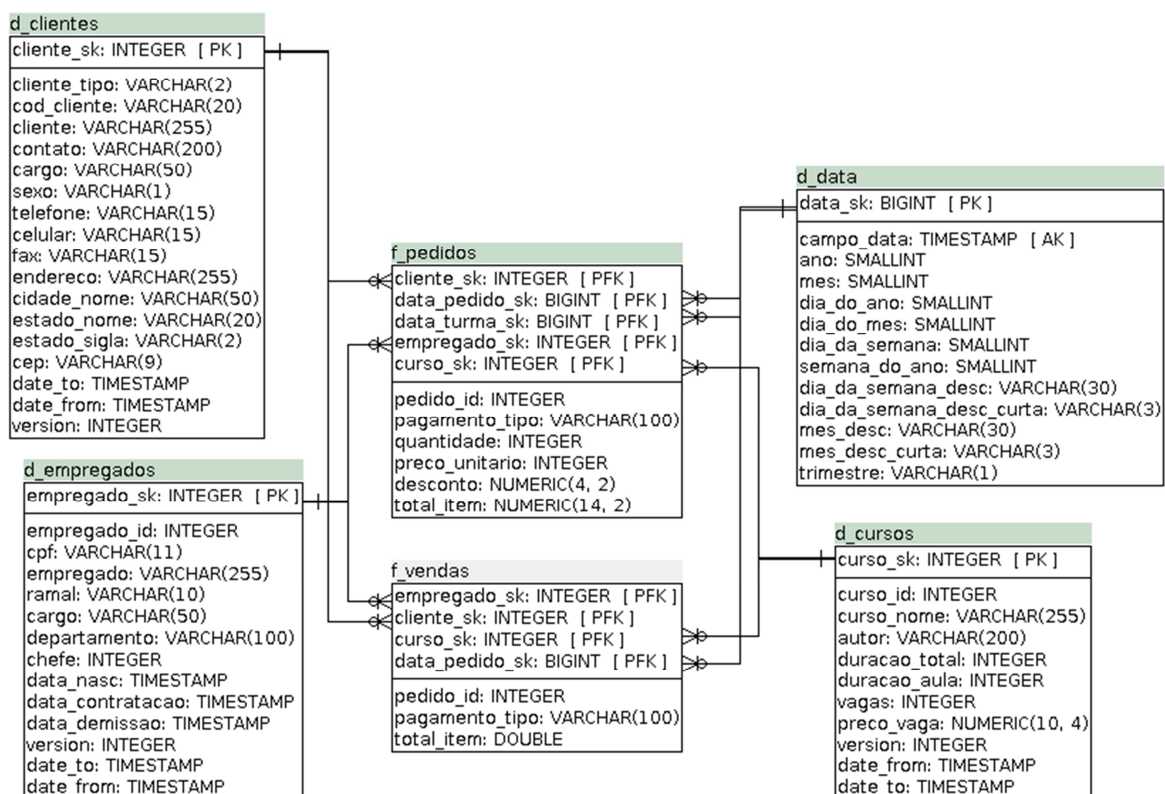


Figura 5 - Diagrama do modelo dimensional de dados da Beltrano S/A  
Fonte: Fabio de Salles<sup>8</sup>

Após isso, inicia-se a fase de construção do relatório no PRD. A figura 6 apresenta a interface inicial da ferramenta.

<sup>8</sup> Disponível em: <<http://geekbi.wordpress.com/2013/09/27/base-de-treinamento-beltrano-sa/>>. Acesso em: 01 nov. 2013.

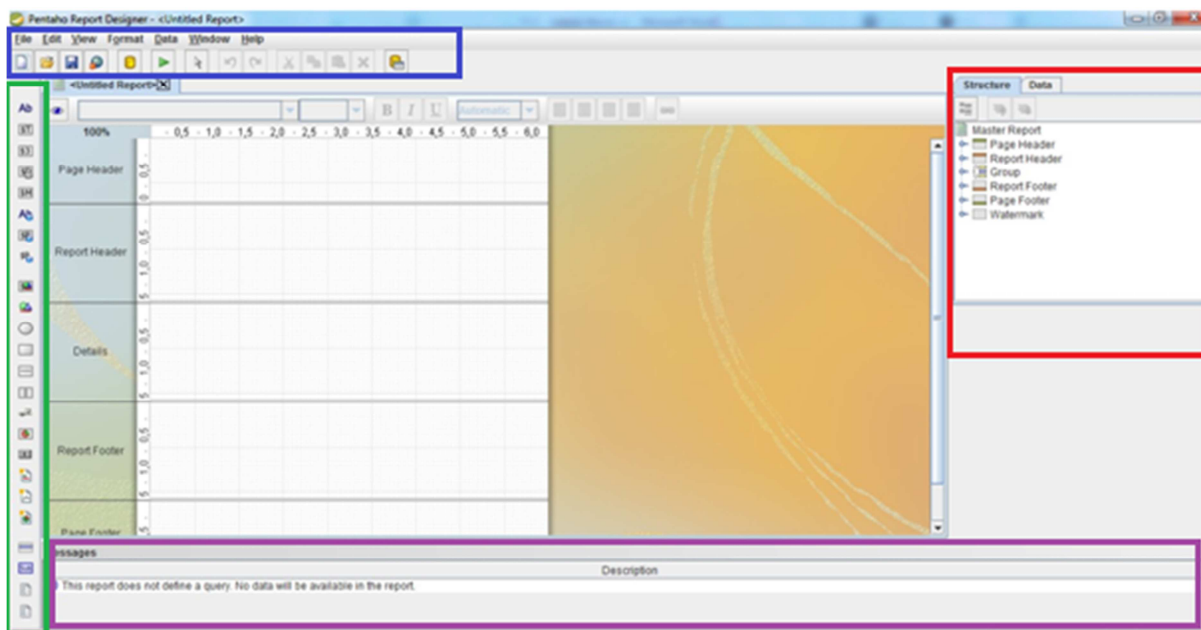


Figura 6 - Interface inicial do PRD

Fonte: Próprio autor

É possível separar essa tela em cinco partes:

A parte destacada em vermelho apresenta a estrutura do relatório na aba *Structure*, com todos os componentes inseridos. Na aba *Data* são exibidas as fontes de dados configuradas.

A parte destacada em verde contém todos os elementos gráficos que podem ser inseridos.

A parte destacada em azul apresenta os menus utilizados para abrir ou criar um novo relatório, publicar o relatório criado, criar novas fontes de dados entre outras coisas.

A parte destacada em roxo é um console de erros e *warnings*.

O centro da tela divide o relatório em cinco partes: cabeçalho da página, cabeçalho do relatório, detalhes, rodapé do relatório e rodapé da página. Todo esse espaço pode ser utilizado para inserção de informações.

#### 4.1. Criação do Relatório

Para apresentar as informações que o gestor necessita serão utilizados os 3 espaços centrais: cabeçalho do relatório, detalhes e rodapé do relatório.

## Cabeçalho do relatório

No cabeçalho normalmente são adicionados o nome do relatório e o logo da empresa. Nessa simulação, ficaria da seguinte forma:



Figura 7 - Cabeçalho do Relatório de Treinamentos  
Fonte: Próprio autor

Para isso, foram adicionados 6 elementos: 1 imagem (logo), 1 campo de texto estático (título) e 4 linhas horizontais (2 acima e 2 abaixo do título e logo). A inserção desses componentes é bastante simples: basta arrastá-los do menu lateral até o local onde deseja que eles fiquem.

No caso da imagem, aparecerá a opção para ir até a pasta em que a mesma está localizada, e então basta selecioná-la. Para o campo de título basta somente arrastar e editar o conteúdo.

## Detalhes

Os detalhes também tem divisão em cabeçalho, corpo e rodapé. No corpo dos detalhes é que acontece a iteração das informações, enquanto no cabeçalho e rodapé elas podem ser dinâmicas, porém não mudarão de estrutura.

Foi criada uma tabela para mostrar a relação de vagas disponibilizadas e as contratadas. Para tanto, a seguinte query foi montada:



```

SELECT
    p.curso_sk as curso_id,
    c.curso_nome as curso_nome,
    count(distinct data_turma_sk) as qtdd_turmas,
    c.vagas as vagas_por_turma,
    trunc(avg(p.quantidade)) as alunos_por_turma,
    c.preco_vaga as preco_vaga
FROM
    d_cursos c, f_pedidos p
WHERE
    p.curso_sk = c.curso_sk
GROUP BY
    p.curso_sk,
    c.curso_nome,
    c.vagas,
    c.preco_vaga;

```

Para montar a tabela foram utilizados apenas campos de texto estáticos (para título das colunas) e campos de texto e de número (para exibir os valores). Observando a figura 8 é possível observar a aparência final da tabela, além da forma em que é escolhido o conteúdo dos campos dinâmicos que apresentam os resultados da query.

The screenshot shows a report builder interface with a table structure and a field selection dropdown. The table has a 'Details Header' section with columns: 'Curso', 'Turmas disponibilizadas', 'Vagas por turma', 'Média de alunos', and 'Valor por vaga'. Below this is a 'Details' section with columns: 'curso\_nome', 'qtdd\_turmas', 'vagas\_por\_turma', 'alunos\_por\_turma', and 'preco\_vaga'. A dropdown menu is open, showing a list of fields with their data types: 'ValorTotalTurmas (java.lang.Number)', 'preco\_vaga (java.math.BigDecimal)', 'alunos\_por\_turma (java.math.BigDecimal)', 'vagas\_por\_turma (java.lang.Integer)', 'qtdd\_turmas (java.lang.Long)', 'curso\_nome (java.lang.String)', 'curso\_id (java.lang.Integer)', and 'report.date (java.util.Date)'.

Details Header		Curso	Turmas disponibilizadas	Vagas por turma	Média de alunos	Valor por vaga
Details		curso_nome	qtdd_turmas	vagas_por_turma	alunos_por_turma	preco_vaga

Figura 8 - Criação de tabela do Relatório de Treinamentos  
Fonte: Próprio Autor

### Rodapé do relatório

Foi criado um gráfico para apresentar a relação entre o valor total arrecadado pela Beltrano S/A e o valor de arrecadado por cada curso, sendo definida a seguinte query para retorno dos dados:

```
SELECT
    curso_sk,
    sum(total_item) as valor
FROM
    f_pedidos
GROUP BY
    curso_sk;
```

Dentro de um mesmo espaço não é possível utilizar queries diferentes para fonte de dados, e então uma será invalidada. Para criar o gráfico foi utilizado o recurso de sub-relatório. Sendo assim, a configuração da query ocorreu nesse novo espaço.

Para obter o valor total de todos os cursos, foi utilizado o recurso de funções agregadoras; nesse caso, a soma de todas as colunas “valor”. A criação da função acontece na aba Structure da seção destacada em vermelho na figura 6.

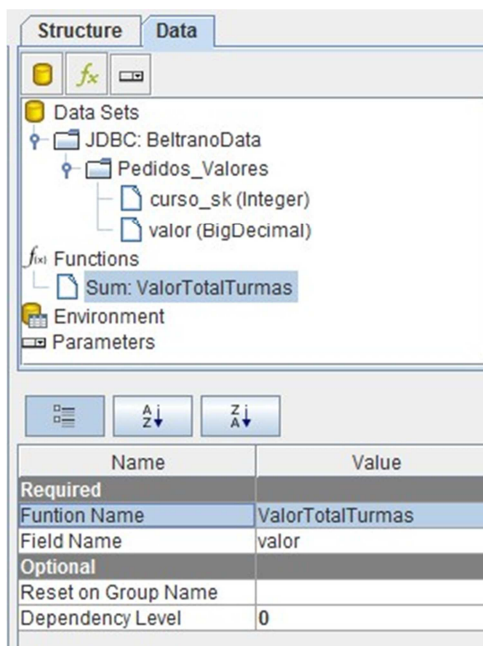


Figura 9 - Função de agregação no Relatório de Treinamentos  
Fonte: Próprio autor

O tipo de gráfico definido foi o de pizza, pois nesse contexto é o que apresenta mais clareza. A Figura 10 mostra a tela de configuração desse gráfico.

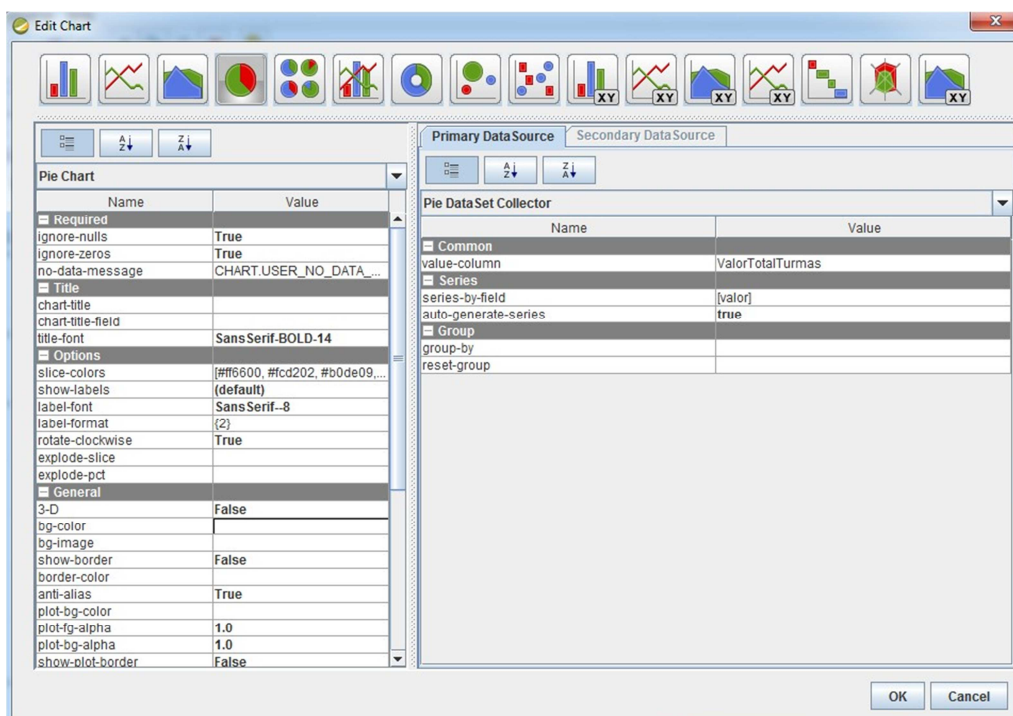


Figura 10 - Criação de gráfico no Relatório de Treinamentos  
Fonte: Próprio autor

Com isso, o relatório é finalizado conforme ilustra a Figura 11.

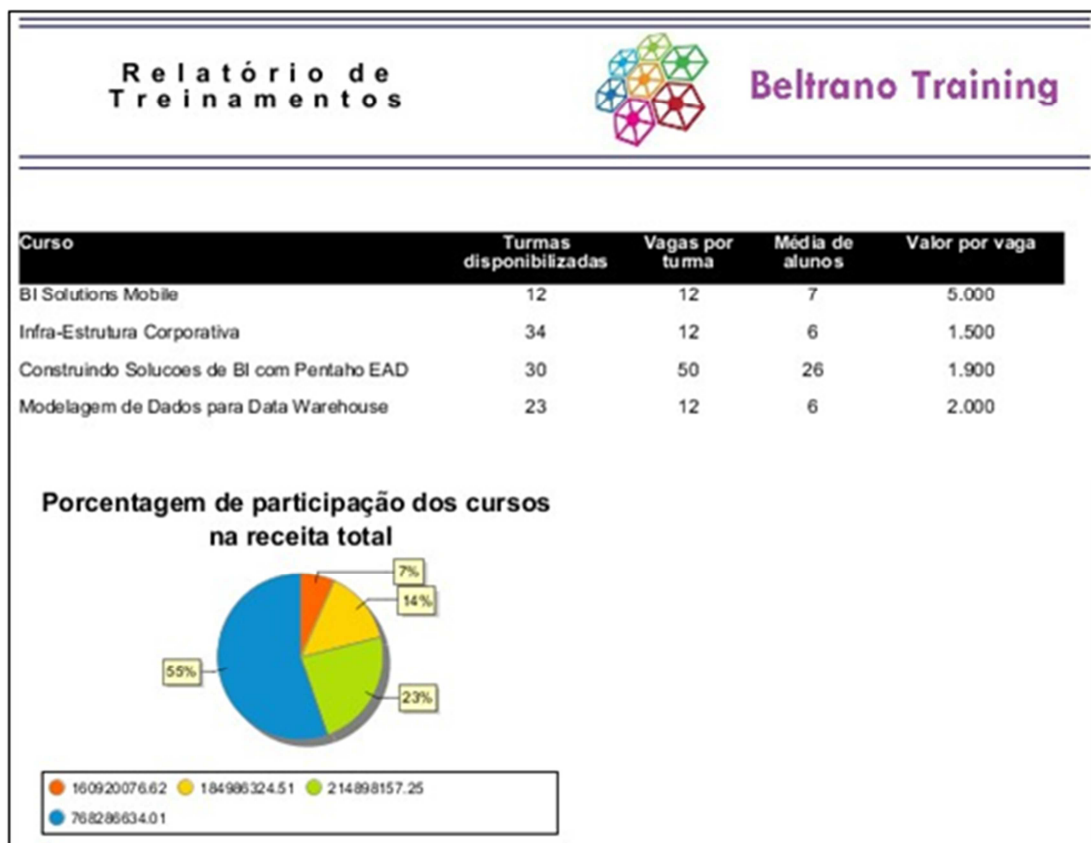
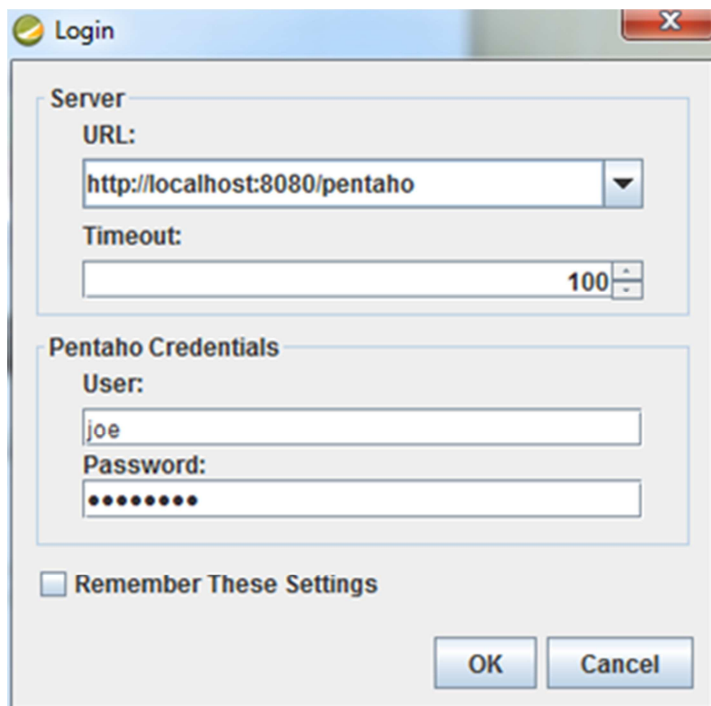


Figura 11 - Relatório de Treinamentos  
Fonte: Próprio autor

#### 4.2. Publicando o relatório no BI Server

Para realizar a publicação de um relatório criado no PRD é preciso que o BI Server esteja rodando. Com isso garantido, a publicação ocorre na interface do PRD.

No menu superior da ferramenta há um ícone para publicação. Ao clicar nesse ícone será necessário configurar o acesso ao BI Server. A Figura 12 mostra quais são as propriedades requeridas.



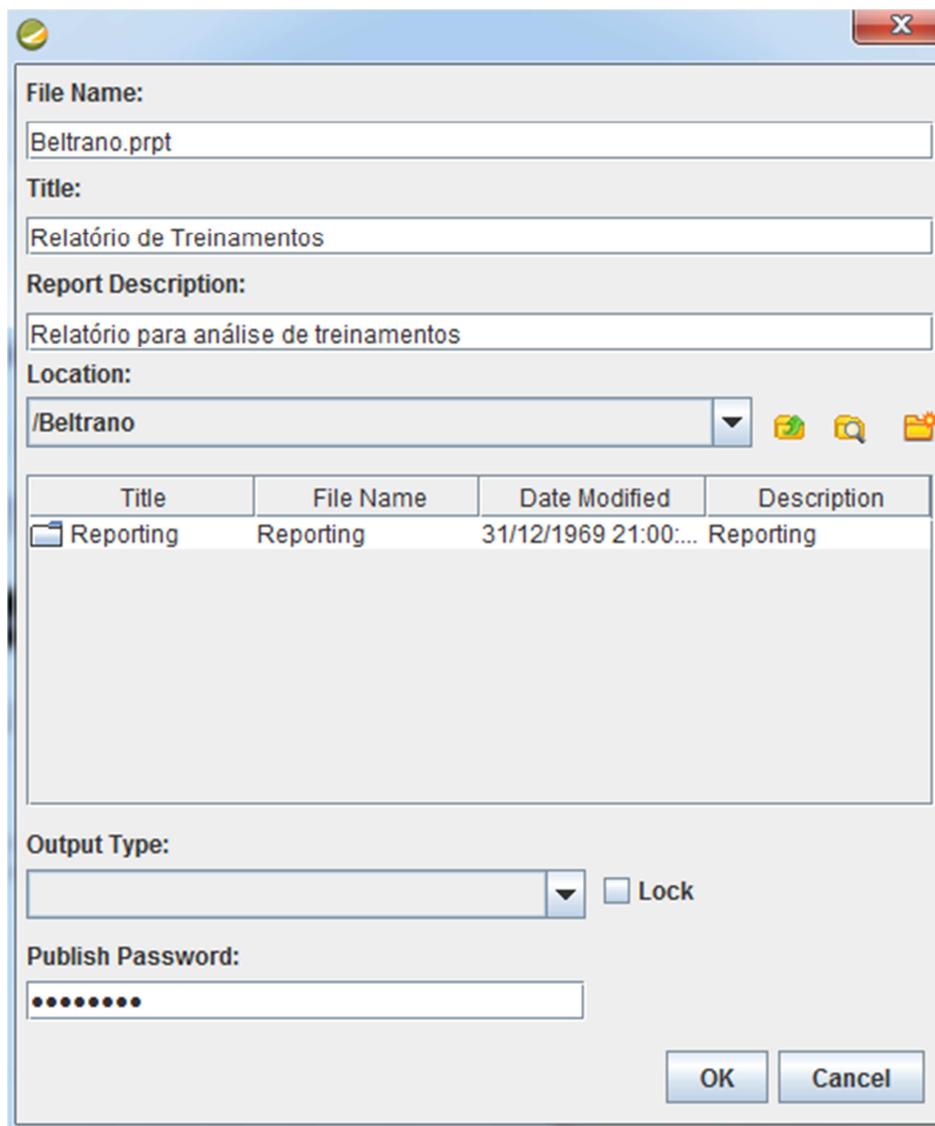
The image shows a 'Login' dialog box with the following fields and controls:

- Server**
  - URL:
  - Timeout:
- Pentaho Credentials**
  - User:
  - Password:
- Remember These Settings
- Buttons: OK, Cancel

Figura 12 - Conexão do PRD com o BI Server  
Fonte: Próprio autor

O usuário utilizado nesse exemplo é um padrão da ferramenta, porém é possível criar ou editar usuário através do console de administração, que vem junto com o pacote do BI Server.

Após isso, os dados serão validados e se o *login* acontecer com sucesso, a próxima etapa é definir nome e local para que o relatório seja armazenado. A figura 13 representa essa etapa.



File Name:  
Beltrano.prpt

Title:  
Relatório de Treinamentos

Report Description:  
Relatório para análise de treinamentos

Location:  
/Beltrano

Title	File Name	Date Modified	Description
Reporting	Reporting	31/12/1969 21:00:...	Reporting

Output Type:  
 Lock

Publish Password:  
.....

OK Cancel

Figura 13 - Publicação do relatório no BI Server  
Fonte: próprio autor

A senha de publicação é definida no arquivo de configuração publisher\_config.xml, presente na estrutura do BI Server em biserver-ce\pentaho-solutions\system.

Ao clicar em OK, o relatório será publicado e poderá ser acessado através da interface do BI Server, onde pode ser acessado por qualquer usuário que tenha autenticação.

#### 4.3. Resultado

A utilização tanto do PRD quanto do BI Server é simples, basta realizar o download na página oficial do Pentaho, configurar as variáveis de ambiente necessárias (CLASSPATH, PATH, JAVA\_HOME e JRE\_HOME). Após isso, basta executar o script que realiza a definição das variáveis utilizadas pelas duas ferramentas, e então iniciá-las.

Antes de se iniciar o uso do PRD é necessário conhecer a(s) base(s) que serão utilizadas, pois é preciso definir as consultas que retornarão os dados a serem exibidos. A ferramenta traz alguns exemplos de relatórios de diversas complexidades, que podem ser utilizados como modelo caso não haja um *design* preferencial já pré-definido.

A interface da ferramenta não é complexa, os elementos visuais são adicionados apenas arrastando-os a partir do menu lateral, porém configurar um gráfico ou um parâmetro pode não ser muito simples, pois podem envolver a criação de funções e tabelas auxiliares que demandam mais planejamento. A publicação no BI Server é um procedimento fácil, desde que exista o conhecimento a respeito do usuário que pode ser utilizado para a publicação (usuários e permissões são definidos no console de administração, que vem junto ao pacote do BI Server, porém é executado separadamente). Também é necessário observar permissões de pastas da ferramenta.

Superando os desafios das duas ferramentas, foi criado um relatório simples, mas que contém informações bastante relevantes. Com esse relatório em mãos, o gestor pode observar que todos os cursos têm muito menos inscritos do que a quantidade total de vagas disponíveis. Além disso, um dos cursos representa mais de 50% da entrada de receita total. A partir dessas informações o gestor tem mais clareza para definir qual a melhor estratégia no momento.

Sendo assim, a ferramenta foi capaz de satisfazer as expectativas iniciais.

## 5. CONSIDERAÇÕES FINAIS

Ao observar as notícias de empresas que caem no mercado por não reconhecer onde estão seus pontos críticos, torna-se necessário refletir a respeito de como as informações são perdidas entre tantos recursos de tecnologia da Informação disponíveis. A partir daí surge a necessidade de olhar para o Business Intelligence.

Esse trabalho apresentou o conceito de Business Intelligence, assim como os componentes presentes em uma solução desse tipo, passando por todas as fases de sua construção. Com isso foi possível constatar que a ideia atrás desse conceito é simples, porém sua execução pode se tornar bastante complexa. É preciso paciência para planejar e estruturar os dados a serem utilizados de forma que seja possível extrair o melhor deles.

Além disso, é necessário investir em tecnologia. Há diversas ferramentas pagas já largamente conhecidas no mercado, porém as licenças chegam a altos preços, dificultando o acesso de várias empresas que ainda não possuem receita suficiente para adquirir esse tipo de recurso. Visando apresentar uma opção mais acessível, especificamente da parte de visualização de dados, foi apresentado o Pentaho.

Utilizando a suíte Pentaho Reporting é possível construir relatórios completos, que devem satisfazer até usuários exigentes. Sua variedade de opções de fontes de dados, seus recursos gráficos, entre outras coisas, atendem as necessidades do mercado, com a vantagem de ser Open Source, ou seja, não gera custos de licença de software. Por não ter uma interface muito complexa, em ambientes de pequeno porte a consultoria pode ser dispensada, porém é bastante indicado que pessoas com conhecimento tanto da tecnologia quanto do negócio façam parte da construção da solução.

Mediante essas informações, através desse trabalho é possível afirmar que o Pentaho é visto como uma ferramenta competitiva de BI, que vem ganhando reconhecimento do mercado (reconhecida pelo Gartner), e apesar de não haver capítulos que apresentassem comparação com outras opções de recursos do mercado que tenham o mesmo objetivo, sua suíte de relatórios é uma solução



indicada para implementação de visualização de dados utilizando BI em empresas de qualquer porte.

Espera-se que esse trabalho sirva como inspiração para mais pesquisas em torno de todos os outros elementos de BI utilizando ferramentas Pentaho.

## 6. REFERÊNCIAS

Ambiente Livre. **Criando um relatório no Pentaho Report Designer com múltiplas fontes de dados usando consultas em fórmulas.**

<<http://www.ambientelivre.com.br/tutoriais-pentaho-bi/272-criando-um-relatorio-no-pentaho-report-designer-com-multiplas-fontes-de-dados-usando-consultas-em-formulas.html>>.

BARBALHO, Patrícia. (2003), **Descubra o Data Warehouse**: produtividade e rapidez, SQL Magazine, Rio de Janeiro, n. 03, p. 34-38.

BATCHERLOR, Marc. **Configuring Publish.**

<<http://wiki.pentaho.com/display/ServerDoc1x/Configuring+Publish>>.

FRANCO, Hilário. **Contabilidade geral**. 23 ed. São Paulo: Atlas, 2009.

INMON, William H. **Como Construir o Data Warehouse**. 2ª ed. Rio de Janeiro: Editora Campus, 1997.

KIMBALL, Ralph. **Data Warehouse Toolkit**: técnicas para construção de Data Warehouses dimensionais. Editora MAKRON, 1998.

MOODY, Glyn. **How Do You Make a Pentaho?**,

<<http://blogs.computerworlduk.com/open-enterprise/2010/05/how-do-you-make-a-pentaho/index.htm>>.

SALLES, Fabio de. **Criar os bancos da Beltrano S/A.**

<[http://www.pentahonapratica.com.br/portal/?page\\_id=132](http://www.pentahonapratica.com.br/portal/?page_id=132)>.

SOUZA, Caio Moreno de. **Curso gratuito online desmistificando BI (Business Intelligence) Open Source (Pentaho e Tecnologias Livres).**

<<http://www.slideshare.net/caiomsouza/curso-gratuito-online-desmistificando-bi-business-intelligence-open-source-pentaho-e-tecnologias-livres>>.

TURBAN, E., Sharda, R., Aronson, J. e King, D. **Business Intelligence**: um enfoque gerencial para a inteligência do negócio. cidade: Artmed,2009, 254 p.