

**O EMPREGO DA FERRAMENTA PENTAHO PARA SOLUÇÕES DE PROBLEMAS  
PARA BUSINESS INTELLIGENCE NO CÂMPUS DE SANTA HELENA**

**Edmar Augusto Yokome<sup>1</sup>**

<sup>1</sup>Docento da UEG-Câmpus Santa Helena de Goiás, e-mail: edmar.yokome@gmail.com

**RESUMO:** A necessidade de analisar dados deixou de ser uma aplicação a mais nos conjuntos de tecnologias de uma empresa para se tornar algo essencial para a tomada de decisões, onde são utilizados: data warehouse, OLAP, mineração de dados dentre outros. Por outro lado devido a grande quantidade de dados que são analisados e a necessidade de uma gama de visões sobre as informações que estes dados podem trazer há a necessidade de uso de ferramentas que auxiliem na análise desses dados, pois caso fosse feita de forma manual poderia levar um tempo muito grande o que poderia talvez tornar a análise defasada ou mesmo haver erros em algum processo que poderia comprometer a integridade dos dados. Dessa forma este artigo tem como objetivo apresentar a ferramenta Pentaho de uma forma geral para a análise de dados, onde a partir dessa ferramenta há um conjunto de módulos para as mais diversas necessidades como migração de dados, análise OLAP, visualização de relatórios e Dashboards dentre outros.

**Palavras-chave:** Ferramentas para Business Intelligence. Análise de Dados. Pentaho.

**Employment Pentaho Tool For Business to Troubleshooting Intelligence on Campus of  
Santa Helena**

**ABSTRACT:** The need to analyze data no longer an application more in sets of technologies from one company to become something essential for decision-making, which are used: data warehouse, OLAP, data mining, among others. On the other hand due to the large amount of data that is analyzed and the need for a range of views on the information that these data can make there is the need to use tools to assist in the analysis of such data, as if it were being done manually it could take a very long time which could perhaps make the lagged analysis or even be errors in any process that could compromise the integrity of the data. Thus this article aims to present the Pentaho tool in general for data analysis, where from this tool is a set of modules for a variety of needs such as data migration, OLAP analysis, view reports and dashboards from others.

**Key-words:** Tools for Business Intelligence. Data analysis. Pentaho.

## **INTRODUÇÃO**

A cada dia que passarmos as corporações dos mais diversos tamanhos vêm utilizando Sistemas de Informação nas mais diversas atividades como uma forma de tornar o processo mais rápido e confiável, dessa forma as bases de dados vêm crescendo de uma forma até então inimaginável nas últimas décadas.

A Ciência da Computação é uma área muito dinâmica e que busca selecionar diversos problemas para a sociedade a sua volta, tendo assim diversas ramificações. Com o tempo pesquisadores fizeram a seguinte pergunta: será que os dados não podem servir para mais coisas além de serem apenas manipulados pelos Sistemas de Informação? Com base nessa pergunta surgiu uma nova ramificação que é chamado de Business Intelligence.

De acordo com Rob e Coronel (2011) Business Intelligence (BI) “é um termo utilizado para descrever um conjunto amplo, coeso e integrado de ferramentas e processos para captar, coletar, integrar, armazenar e analisar dados para a geração e apresentação de informações e a apresentação de informações que dêem suporte à tomada de decisões de negócio”.

De acordo com Santos e Ramos (2009) os sistemas de Business Intelligence estão tradicionalmente associados a três tecnologias: Data Warehouses, On-Line Analytical Processing e Data Mining. Com as novas necessidades que vem surgindo a cada ano novas tecnologias vem sendo desenvolvida entre elas podemos destacar o Big Data que cada vez mais está tendo uma grande variedade de aplicações.

De acordo com Machado (2010) Data Warehouse representa uma grande base de dados capaz de integrar, de forma concisa e confiável, as informações de interesse para a empresa, que se encontram espalhadas pelos sistemas operacionais e em fontes externas, para posterior utilização nos sistemas de apoio à decisão.

De acordo com Santos e Ramos (2009) OLAP é uma tecnologia utilizada para explorar Data Warehouses onde é criado cubos para analisar a informação sob diferentes perspectivas.

De acordo com TAN, STEINBACH e KUMAR (2009) é o processo de descoberta automática de informações úteis em grandes depósitos de dados. Para que isso ocorra são aplicados algoritmos que permitem que dados possam se associar a outros revelando padrões que por meio de técnicas convencionais não seria possível estar descobrindo.

**10ª Jornada Acadêmica da Jornada da UEG**  
**“Integrando saberes e construindo conhecimento”**  
**10 a 12 de Novembro de 2016**  
**UEG - Câmpus Santa Helena de Goiás, GO**

Na computação é bem comum o uso de ferramentas para tornar o trabalho mais rápido e confiável e na área de Business Intelligence não poderia ser diferente, no mercado há um conjunto de ferramentas que podem ser utilizadas entre elas estão: Oracle Business Intelligence Standard, Microsoft Business Intelligence, SAP Business Intelligence, SAS Business Intelligence, IBM Cognos, Pentaho Community, SpagoBI dentre outras.

De acordo com Oliveira (2015) a plataforma Pentaho, conhecida como Pentaho Open BI, é composta de aplicativos open source para criação de soluções de Business Intelligence. Entre os módulos destacam-se: Pentaho BI Plataform, Pentaho Data Integration (Kettle), Analysis View (Modrian) e Pentaho Reporting (Reporting).

Neste artigo o objetivo é apresentar a ferramenta Pentaho Community no uso do Câmpus de Santa Helena onde a ferramenta está sendo utilizados para o desenvolvimento de Trabalho de Curso, Projetos de Pesquisa e em disciplinas com conteúdo livre para ser aplicado em determinado ano. O próximo tópico trata especificamente de mostrar como esta ferramenta vem sendo utilizada no Câmpus de Santa Helena.

## **DESENVOLVIMENTO**

Neste tópico são apresentados alguns usos da Ferramenta Pentaho no Câmpus de Santa Helena, assim será apresentado o uso em um projeto de pesquisa e também na disciplina de Novas Aplicações em Sistemas de Informação (NASI).

A primeira aplicação foi e está sendo em um projeto de pesquisa intitulado Implantação de uma base de dados para Aquisição, tratamento edisponibilização de dados meteorológicos em um sistema de informação WEB neste projeto foi criado uma estação meteorológica em Arduinoe os dados que são gerados são XML contendo dados como: temperatura, direção do vento, quantidade de chuva, velocidade do vento e umidade. No Câmpus há outra estação modelo ITWH-1080 que gera um arquivo .xls com os mesmos dados citados acima. Estes dados ficam armazenados em uma unidade de memória de seus respectivos dispositivos.

Para que os dados sejam processados neste projeto foi adotado um banco de dados relacional onde a ideia é armazenar os dados para posteriormente processá-los, porém para que isso aconteça é necessária uma aplicação que transfira os arquivos para uma base de dados para isso é utilizadaa Ferramenta Pentaho. A Figura 1 apresenta como foi utilizada a ferramenta para pegar um arquivo em .xls e migrar para uma base de dados relacional.

**10ª Jornada Acadêmica da Jornada da UEG**  
**“Integrando saberes e construindo conhecimento”**  
**10 a 12 de Novembro de 2016**  
**UEG - Câmpus Santa Helena de Goiás, GO**



Figura 1: Migração dos dados do tipo XML para o MYSQL

Ao analisar a Figura 1 é possível identificar dois componentes o primeiro chamado Dados XML, onde é configurado este componente com o endereço do arquivo e inserido na listagem do arquivo, após isso os dados dos arquivos estão prontos para realizar tarefas posteriores. No segundo componente é configurado a tabela de uma base de dados para receber os dados do primeiro componente, para isso é necessário o driver do banco de dados, a criação de um banco de dados com suas respectivas tabelas (onde os campos das tabelas devem ter os mesmos campos do arquivo XML). Por fim tem uma linha que liga um componente ao outro onde os dados do primeiro objeto são passados para o segundo objeto. Acima dos objetivos tem uma seta verde onde quando tudo estiver pronto então é acionado e então ocorre o processo de migração de dados.

Na Figura 2 é apresentada parte dos dados armazenados na base de dados criados no Sistema de Gerenciamento de Dados (SGBD) MYSQL.

	Data de coleta	Hora Coleta	Temperatura	Umidade Relativa	Ponto de Condensação
▶	02/03/2010	17:02:18	31.5	45.6	18.3
	02/03/2010	17:02:18	31.5	45.6	18.3
	02/03/2010	17:03:18	31.6	49.3	19.6
	02/03/2010	17:04:18	31.5	49.3	19.5
	02/03/2010	17:05:18	31.4	50.7	19.9
	02/03/2010	17:06:18	31.4	51.0	20.0
	02/03/2010	17:07:18	31.3	51.4	20.0

Figura 2: Visão de Parte dos Dados Armazenados

A partir desses dados armazenados dê início no projeto foram gerados apenas relatórios bem simples para isso foi utilizado outra suíte do Pentaho chamada PentahoReporting a intenção neste projeto de pesquisa é estar cada vez mais estar aperfeiçoando os relatórios e também gráficos, assim foi proposto, aprovado e está em execução um projeto de pesquisa para aumentar o leque de tratamento dos dados.

**10ª Jornada Acadêmica da Jornada da UEG**  
**“Integrando saberes e construindo conhecimento”**  
**10 a 12 de Novembro de 2016**  
**UEG - Câmpus Santa Helena de Goiás, GO**

No projeto de pesquisa o uso dessa ferramenta está mais vinculado ao processo de migração de dados e geração de relatórios aproveitando apenas uma parte das grandes funcionalidades que a ferramenta fornece. Para um melhor aproveitamento da ferramenta no primeiro semestre de 2016 esta ferramenta foi apresentada aos acadêmicos do terceiro ano do curso de Sistemas de Informação da UEG-Câmpus de Santa Helena a seguir serão relatados como esta ferramenta foi utilizada.

No curso de Sistemas de Informação tem uma disciplina chamada NASI onde o objetivo é que a cada ano um professor pegue a disciplina e aplique alguma tecnologia que está em evidência no mercado, sendo assim no primeiro semestre de 2016 a proposta era que fosse introduzido o conceito de Business Intelligence aos acadêmicos do curso.

Inicialmente foi estudada o Data Warehouse com foco na modelagem, criação de migração de dados, para isso foi utilizado o conceito de modelagem Multidimensional baseado no modelo em estrela e floco de neves, ainda foi utilizado o Sistema Gerenciador de Banco de dados MYSQL e algumas bases de dados vista na disciplina de Banco de Dados. A partir das bases de dados existentes foram modelados os Data Warehouse e também compilado os códigos fonte gerado com a modelagem e por fim a criação. Após a execução dessas atividades foi necessário arrumar uma forma de transferir uma base de dados para a outra, para que isso fosse possível foi utilizado o Pentaho, porém antes disso foram feitos estudos na ferramenta como um todo. A Figura 3 apresenta um exercício feito nesta ferramenta.

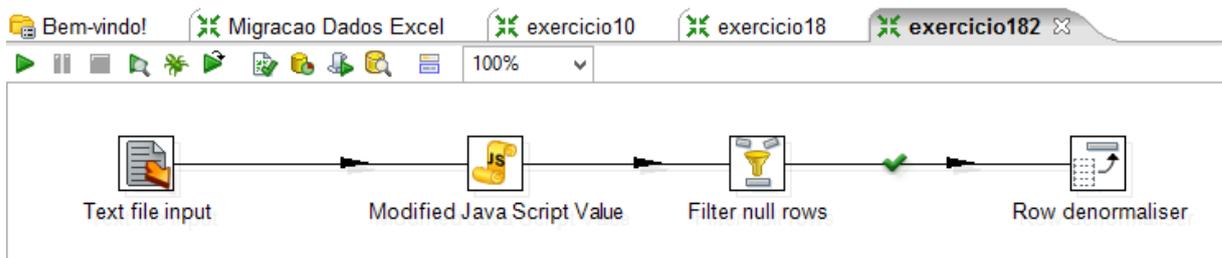


Figura 3: Exercício feito no Pentaho

Este exercício consiste em obter um texto qualquer e retirar os espaços em branco desse texto. Assim foram inseridos quatro objetos, o primeiro objeto consiste em obter um determinado texto, o segundo objeto consiste em permitir a criação de scripts na Linguagem de Programação Java Script para identificar os espaços em branco, o terceiro elemento consiste em filtrar os espaços em branco deixando apenas os demais caracteres e por fim o quarto elemento recebe o novo texto sem os espaços em branco.

Após alguns exercícios como estes foi possível então ter uma noção de como a ferramenta funciona e então fazer o processo de migração de dados de uma base de dados relacional para um Data Warehouse criado conforme é apresentado na Figura 4.

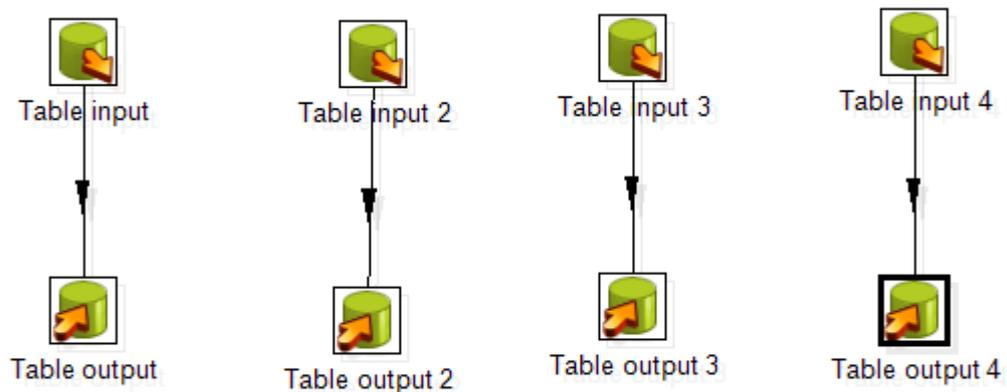


Figura 4: Processo de Migração de Dados

Na Figura 4 é possível apresentar como ocorre o processo de migração de dados, onde a idéia é que tanto na base relacional como no Data Warehouse possuem tabelas e foram criados a partir de um SGBD, a partir do primeiro componente chamado Table Input são configurados o driver do SGBD e os parâmetros de configuração (isso no banco de dados relacional onde os dados estão) e no segundo objeto chamado Table Output é configurado a tabela do Data Warehouse para receber os dados. Depois que estiver tudo configurado base apertar o botão de executar e então os dados que estavam na base de dados relacional irá para o Data Warehouse.

Estava programado ainda na disciplina gerar relatórios a partir dos dados armazenados no Data Warehouse, fazer análise em cubos utilizando o OLAP e aplicar mineração de dados, porém não foi possível devido a mudança de professor na disciplina.

## CONCLUSÕES

Este artigo teve como objetivo apresentar a ferramenta Pentaho como uma opção para aplicações de Business Intelligence, onde as principais características da ferramenta é ter uma versão opensource e outra paga, porém a versão opensourcedeixa apenas de oferecer alguns recursos de relatórios e gráficos que a outra versão possui. Além disso, a ferramenta possui uma variedade de módulos que vem atender a diversas necessidades e de forma separada permitindo assim uma maior flexibilidade no uso da ferramenta.

**10ª Jornada Acadêmica da Jornada da UEG**  
**“Integrando saberes e construindo conhecimento”**  
**10 a 12 de Novembro de 2016**  
**UEG - Câmpus Santa Helena de Goiás, GO**

Na disciplina de Novas Aplicações em Sistemas de Informação no primeiro semestre de 2016 do curso de Sistemas de Informação o professor da então disciplina tratou do assunto Business Intelligence e utilizou esta ferramenta em suas aulas. Foi possível comprovar que a ferramenta é de fácil utilização onde são inseridos e configurados objetos de acordo com o que se deseja realizar.

Outros usos dessa ferramenta no Câmpus de Santa Helena estão ligados a um projeto de pesquisa onde são gerados arquivos em Excel e XML e então os dados são armazenados em uma base de dados relacional. E também um aluno que está fazendo Trabalho de Curso tem trabalhado com a ferramenta aplicando em uma empresa.

Para finalizar a ferramenta Pentaho tem atendidos bem os projetos que vem sendo desenvolvido no Câmpus, mostrando ser uma ferramenta versátil e flexível para um conjunto de atividades de BI que necessitam ser realizadas.

## **REFERÊNCIAS**

ROB, Peter; CORONEL, Carlos. **Sistemas de Banco de Dados – Projeto, Implementação e Administração**. Editora: Cengage Learning. São Paulo, 2011.

MACHADO, Felipe Nery Rodrigues. **Tecnologia e Projeto de Data Warehouse**. Editora: Érica. São Paulo, 2010.

SANTOS, Maribel Yasmina; RAMOS, Isabel. **Business Intelligence – Tecnologias da Informação na Gestão de Conhecimento**. Editora: FCA. Lisboa, 2009.

TAN, Pang-Ning; STEINBACH, Michael; KUMAR, Vipin. **Introdução Ao Data Mining**. Editora: Ciência Moderna. Rio de Janeiro, 2009.