

Análise comparativa dos pacotes de ferramentas open source para data warehouse

Resumo

A pesquisa tem como objetivo mostrar uma comparação dos pacotes de ferramentas open source para data warehouse (DW). Também conhecidas como ferramentas de BI (Business Intelligence), elas são cada vez mais utilizadas pelos profissionais responsáveis pela gestão das organizações. A comparação da suíte de ferramentas irá verificar qual é a mais completa, qual abrange o maior número de aplicações. Em seguida detalharemos o uso e a funcionalidade de cada ferramenta do pacote escolhido. Essas informações devem ajudar os tomadores de decisões das empresas, oferecendo uma solução livre de custos de licença e os profissionais que atuam na área de banco de dados que não conhecem ou não tem tempo para pesquisar as novidades em relação às ferramentas open source do mercado

Palavras chave: data warehouse, bi, ferramentas data warehouse.

1. Introdução

Atualmente a distribuição das empresas em diversas regiões geográficas gera um enorme volume de dados e informações, tratar esses dados filtrando informações relevantes, é um dos motivos que estimula o uso de arquiteturas de data warehouse. Os bancos de dados têm sido de grande valia nas aplicações de business intelligence para análise de dados e apoio ao processo de tomada de decisões e são cada vez mais utilizados por empresas de qualquer porte ou ramo. O DW, componente principal do data warehouse, consiste em um banco de dados especial, organizado para armazenar uma coleção de dados integrados, orientados por assunto e variáveis, que serve para dar suporte às altas gerências das empresas através de análise de tendências históricas dos seus produtos, funcionários e clientes, aumentando, com isso, a competitividade da mesma. Ter conhecimento desse universo é de extrema importância para quem se interessa pela área de tecnologia e banco de dados. Progressivamente as empresas vêm reconhecendo a importância da informação para seus negócios e precisam de profissionais que tenham conhecimento nesse tipo de estrutura. O objetivo da pesquisa é fazer um estudo comparativo entre os pacotes de ferramentas open source para data warehouse, descrevendo as ferramentas e suas aplicações. Atualmente o mercado oferece várias ferramentas e por isso surge a pergunta:

Qual o mais completo e eficiente pacote de ferramentas open source para data warehouse disponível hoje no mercado?

O objetivo geral da pesquisa é mostrar os pacotes de ferramentas open source para DW e detalhar as ferramentas existentes no pacote que abrange o maior número de atividades, ou seja, o mais completo. Percebe-se assim que a aplicação de data warehouse está baseada em várias idéias dentre elas: integrar fontes de dados heterogêneas e elaborar

relatórios mais precisos para os tomadores de decisão. O pacote de ferramentas que atender da melhor forma essas expectativas será considerado o mais eficiente. Para isso é necessário:

- Definir data warehouse, sua arquitetura, modelagem, armazenamento e extração de dados através das ferramentas OLAP (Online Analytical Processing) e OLTP (Online Transaction Processing).[10],[11]
- Identificar os principais pacotes de ferramentas para data warehouses, comparar as ferramentas que eles oferecem e suas aplicações.
- Detalhar cada ferramenta da suíte que foi escolhida como a melhor. Mostrar instalação, configuração e o uso de cada ferramenta em particular.

Esse documento está organizado da seguinte maneira. Na próxima sessão é apresentado o estado da arte, na seção 3 a metodologia ou revisão bibliográfica e em seguida o cronograma do projeto e as referências.

2. Estudo da arte

A pesquisa bibliográfica mostrou que apesar de existir vários artigos e teses a respeito do tema poucos exploram os softwares que são utilizados no data warehouse e que produzem ferramentas de trabalho tão eficazes.

É perceptível o início da abordagem pelos pesquisadores ao tema no início dos anos 80 período de surgimento dos primeiros bancos de dados integrados. No decorrer do período até os dias atuais ocorreu uma grande evolução dos data warehouses e seus possíveis campos de atuação.[3]

Dentre as técnicas de um data warehouse destacam-se: redes neurais, árvores de decisão e regra de indução. Os métodos e implementações devem estar aliados ao plano de negócio da empresa e suas estruturas de dados devem ser bem analisadas para garantir o sucesso da implementação. [15]

Em [7] é feito uma comparação entre as ferramentas BI livres e as pagas. Várias ferramentas foram pré-selecionadas, fatores como funcionalidades, componentes, custo e amadurecimento foram analisados. A ferramenta livre selecionada foi a Pentaho, enquanto que a paga escolhida foi a da Microsoft. Um estudo de caso mostra a utilização das duas ferramentas. [4] apresenta em sua pesquisa uma análise comparativa de ferramentas Olap. Segundo [18] Este artigo revê o conceito de Business Intelligence e fornece um levantamento, do ponto de vista global, do quadro técnico BI, processos e soluções empresariais. Além disso, as conclusões apontam as possíveis razões para as dificuldades de implantação ampla de BI corporativo, e as propostas de construção de um melhor sistema de BI. [17] apresenta uma comparação das ferramentas OLAP com relação a segurança e integridade dos dados. [5] apresenta em seu estudo os processos envolvidos na construção de um sistema que faz uso da tecnologia de DW utilizando ferramentas open source. Alguns estudos abordam as ferramentas open source OLAP que trabalham com análise de dados on line, outros comparam as ferramentas ELT

(extração, transformação e carga). Contudo as suítes de ferramentas existentes são bem mais abrangentes.

[9],[14],[8],[16],[2],[6],[1],[13],[12]

3. Metodologia

Em andamento

4. Cronograma

Em andamento

5. Desenvolvimento

Em andamento

5. Conclusão

Em andamento

6. Referências

[1] Agata. Acessado em 15/05/2010.

[2] Apatar. Acessado em 15/05/2010.

[3] C. A. F. BISPO. *Uma análise da nova geração de sistemas de apoio a decisão*. PhD thesis, Departamento de Informática, Universidade de São Paulo, 1998.

[4] P. C. S.; MONTEIRO R. R.; SOARES V. J. CAVALCANTI, M. C.; OLIVEIRA. *Análise comparativa de ferramentas olap*. Projeto de pesquisa-UFRJ, 1998.

[5] B.O. Dias. *Projeto de data mart utilizando ferramentas open source*. Projeto de pesquisa-UFPA.

[6] Eclipse. Acessado em 15/05/2010.

[7] R.; Vieira V.; Guimarães C.; Carvalho J. Ferreira, M; Silva. *Um estudo de caso com análise comparativa entre ferramentas de bi livre e proprietária*. Projeto de pesquisa - UFPA.

[8] Jasper. Acessado em 15/05/2010.

[9] keydata. *Data warehousing*. Acessado em 30/03/2011.

[10] Margy KIMBALL, Ralph ; ROSS. *The Data Warehouse Toolkit*. John Wiley and Sons, 2001.

[11] Margy; THORNTHWAITE Warren KIMBALL, Ralph; ROSS. *The Data Warehouse: Lifecycle Toolkit*. John Wiley and Sons, 1998.

[12] Microstrategy. Acessado em 15/05/2010.

[13] Palo. Acessado em 15/05/2010.

[14] Pentaho. *Pentaho bi*. Acessado em 15/05/2010.

[15] Jorge Vaz de Oliveira Sá. *Metodologias de sistemas Data Warehouse*. PhD thesis, Departamento de tecnologia, Universidade do Minho, 2009.

[16] Talend. Acessado em 15/05/2010.

[17] Günther Pernul Torsten Priebe. *Towards olap security design survey and research issues*. Projeto de pesquisa-University of Essen.

[18] Li Zeng; Lida Xu; Zhongzhi Shi; Maoguang Wang; Wenjuan Wu;. *Techniques, process, and enterprise solutions of business intelligence*. *Systems, Man and Cybernetics*, 6:4722-4726, 2006.