AVALIAÇÃO DE FERRAMENTAS OLAP PARA MANIPULAÇÃO E EXTRAÇÃO DE DADOS NA AGRICULTURA DE PRECISÃO

Emerson Pinheiro Marinho da Silva (UEPG)
Maria Salete Marcon Gomes Vaz (Orientadora), e-mail: salete@uepg.br

Universidade Estadual de Ponta Grossa/Setor de Ciências Agrárias e de Tecnologia/Departamento de informática

Metodologia e Técnicas da Computação/Banco de Dados

Palavras-chave: Banco de Dados, *Data Warehouse*, Modelagem Multidimensional, OLAP.

Resumo

Nesta pesquisa foram estudadas ferramentas OLAP, bem como os conceitos envolvendo *data warehouse* e modelagem multidimensional, visando a aplicação dessas ferramentas na agricultura de precisão. Foram levantadas ferramentas OLAP de modo a avaliar e comparar suas funcionalidades. Além disso, foi avaliado o cenário de integração da modelagem de *data warehouse* agrícola espacial, auxiliando a tomada de decisão para o planejamento e gestão do manejo das práticas da agricultura de precisão, com implementação de bases de dados multidimensionais. Como resultado, houve geração de conhecimento sobre ferramentas OLAP, em bases de dados multidimensionais, como uma ferramenta para auxiliar os agricultores no gerenciamento de suas lavouras.

Introdução

Atualmente, são produzidas enormes quantidades de dados, provenientes de diversos dispositivos, como computadores, smartphones, dispositivos embarcados, etc. Visando seu aproveitamento, foram criadas ferramentas capazes de analisa-los, a fim de extrair informação relevante, utilizada principalmente no âmbito das tomadas de decisões estratégicas. Uma ferramenta de manipulação de *data warehouse* está nesse contexto.

O setor agrícola se beneficia do desenvolvimento tecnológico, na utilização de ferramentas para, em geral, melhorar a produção. A agricultura de precisão é uma tecnologia e pode ser utilizada para o manejo do solo, dos insumos e das culturas, de modo adequado e considerando as variações espaciais e temporais dos fatores que afetam a produtividade (ANTUNIASSI; BAIO; SHARP, 2017).

Com a tecnologia na agricultura, este campo passou a gerar dados que, se bem aproveitados podem gerar informações estratégicas para o manejo das lavouras, melhorando a sua produtividade. A tecnologia de *data warehouse*, associada à ferramentas que possibilitem a extração dessas informações, serve de auxilio nesta tarefa.

Revisão de literatura

O *data warehouse* é uma arquitetura que define um armazém de dados, referentes às regras de negócio da empresa, integrando e consolidando informações de fontes

internas e externas, e sumarizando, filtrando e limpando esses dados, preparando-os para a análise e suporte a decisão (MACHADO, 2009).

Os *data warehouses* possuem características, tais como orientação por assunto, onde as informações são agrupadas por áreas de interesse da empresa, e variação de tempo, onde, diferente dos bancos de dados transacionais, os *data warehouses* possuem espaços de tempos significativamente maiores, situando-se entre cinco e dez anos. Além disso, as informações são não voláteis, visto que os dados não se perdem, já que as únicas operações executadas pelo *data warehouse* são a carga e a consulta de dados (MACHADO, 2009).

As ferramentas utilizadas para extrair informações de um *data warehouse* são chamadas de Ferramentas OLAP - *Online Analytical Processing*, aplicações às quais os usuários finais têm acesso para extrair os dados de suas bases e construir relatórios capazes de responder às suas questões gerenciais (MACHADO, 2009). O processo que envolve a extração, a transformação e o carregamento dos dados para o *data warehouse* é chamado de ETL – do inglês *Extract, Transform, Load*.

Em geral, espera-se que as ferramentas OLAP promovam o acesso *ad hoc* à informação, suportem os complexos requerimentos de análise de tomadores de decisão, permitam a análise dos dados de diferentes perspectivas, além de permitir análises complexas em um grande volume de dados (ABDULLAH, 2009).

Desse modo, estas ferramentas executam um conjunto de operações, extraindo informações úteis aos usuários finais. Machado (2009) descreve estas operações, resumidamente compreendem o processo de aumento e diminuição de detalhes de dados (Drill up/down); mudança dos níveis de informação (Drill across); mudança da dimensão (Drill throught); filtragem de informações e mudança na perspectiva dos dados (Slice e Dice).

Para realizar estas operações, tratando-se de *data warehouse*, utiliza-se a modelagem multidimensional. Nela, os dados podem ser agrupados em matrizes multidimensionais chamadas cubos de dados (*data cubes*), no qual suas dimensões correspondem à aspectos comuns de negócio como, por exemplo, produto, período fiscal, etc. Além disso, este modelo possui uma performance de consulta muito melhor que o modelo relacional (*star schema*) (ELMASRI, NAVATHE, 6ª ed., 2011).

Machado (2009) destaca 3 elementos básicos do modelo multidimensional:
Fato – Implementado nas tabelas de fatos (*fact tables*), representa a evolução dos negócios do dia-a-dia de uma organização; Dimensões – São os assuntos de negócio e podem ser hierarquicamente divididas; Medidas (Variáveis) – Indicador numérico dos negócios relativos às dimensões que participam desse fato.
A figura 1 ilustra este conceito no contexto da agricultura.

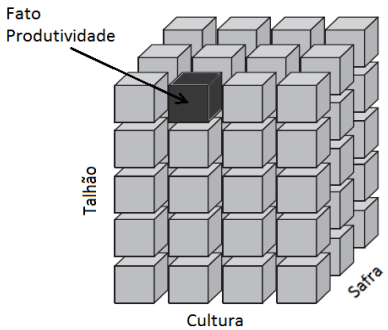


Figura 1 – Modelagem multidimensional da área agrícola.
Fonte: Adaptado de: Petroski, L. P., p. 23, 2017

Podem ser encontrados diversos exemplos do uso de *data warehouse* e ferramentas OLAP ligados à agricultura e à pecuária. Rai *et al.* (2008) descrevem em seu artigo o projeto e o desenvolvimento de um data mart para recursos animais, no qual fazem uso de tais tecnologias para desenvolvimento de um sistema de suporte a decisão.
Já no campo de agricultura de precisão especificamente, é possível encontrar exemplos como o mostrado por Abdullah (2009), no qual é desenvolvido um sistema de suporte de decisão para auxiliar agricultores no controle da incidência de Cochonilha em colheitas de algodão.

Para o setor da agricultura, um aspecto essencial a ser levado em conta são os dados espaciais, bem como suas características de modelagem e manipulação. Assim, pode-se aliar as ferramentas OLAP com Sistemas de Informação Geográfica (SIG), a fim de atender este requisito.

Depois da revisão bibliográfica, avaliou-se ferramentas OLAP para fins de comparação. Foram levantadas as seguintes ferramentas: Knowage (antes Spago BI), Cubes Viewer e Superset.

Resultados e Discussão

A primeira ferramenta avaliada foi o Knowage, uma solução composta por diversos serviços, compreendendo desde ferramentas de Big Data até análises preditivas de indicadores.

Knowage se mostra como uma ferramenta robusta no tratamento de dados. As funcionalidades relacionadas à modelagem dimensional e OLAP, vão desde a possibilidade da criação dos schemas, até sua visualização, na qual o usuário pode arrastar e soltar as dimensões do cubo, cruzando diferentes informações. Existem ainda opções mais complexas, como a utilização de queries MDX para consultas.

Destacam-se ainda, as possibilidades de se trabalhar com modelagem espacial, visualizando mapas, aplicando filtros em dados espaciais, etc., e a criação de KPI (Key Performance Indicator), permitindo o acompanhamento de indicadores de performance do negócio. Ainda estão presentes outros modelos de visualização dos dados, como gráficos, histogramas, resumos, etc. É possível também, ao administrador, personalizar diferentes partes da ferramenta.
A segunda ferramenta analisada foi o Superset. Mais leve, simples e rápida comparada à anterior, conta com diversas funcionalidades presentes na primeira. Apresenta a possibilidade de tratamento de dados espaciais, com mapas, além de possuir diversos tipos de visualizações, como gráficos, tabelas, etc. É possível ainda personalizar e criar dashboards (painéis) de acordo com as necessidades dos usuários.

A ferramenta apresenta também uma sessão em que é possível criar consultas SQL, armazenando-as para posterior utilização. O Superset porém, diferente do Knowage, não apresenta a possibilidade da criação dos schemas, mas a utilização das tabelas Pivot (cubos de dados OLAP) funciona de maneira parecida. Outra desvantagem é a documentação, que não é tão completa como a anterior. Apesar de não ser tão completa quanto à anterior, mostra-se eficaz quanto ao seu uso como ferramenta OLAP.

Por último, analisou-se o Cubes Viewer. Esta ferramenta é específica para operações OLAP, porém é muito simples. Além disso, não apresenta modelagem

espacial com mapas nem diversidade de personalização e visualizações. Ou seja, é a ferramenta mais limitada dentre as 3 analisadas.
A tabela 1 resume de maneira simplificada alguns pontos observados:

Tabela 1 – Comparação das ferramentas OLAP analisadas

Ferramenta	Suporte a modelagem espacial	Complexidade	Interface	Licença
Knowage	Sim	Alta	Boa	Open Source para a versão community
Superset	Sim	Média	Boa	Open Source
Cubes Viewer	Não	Média	Regular	Open Source

Conclusões

Como conclusão, destaca-se que a utilização de técnicas de *data warehouse* atreladas ao setor agropecuário promove o avanço do setor, não somente na coleta de informações, mas na utilização para a melhoria da produtividade. Porém, deve-se ter cautela na implementação, visto que há, além do custo de implementação, a necessidade de mão de obra especializada, e ainda é necessária uma avaliação prévia da ferramenta que se pretende utilizar. Deste trabalho tem-se que Knowage e Superset são boas opções. O destaque negativo foi a ferramenta Cubes Viewer que, à primeira vista, não apresenta a opção de modelagem espacial, essencial para o campo da agricultura.

Referências

ANTUNIASSI, U. R.; BAIO, F. H. R.; SHARP, T. C. **AGRICULTURA DE PRECISÃO**. Disponível em < https://www.agencia.cnptia.embrapa.br/Repositorio/1622_000fkl0f2ta02wyiv80sq98yqf7fpgf0.pdf> Acesso em: 9 mar 2017

CUBESVIEWER. Disponível em <https://github.com/jjmontesl/cubesviewer>. Acesso em 15 nov 2016.

MACHADO, Felipe Nery Rodrigues. **TECNOLOGIA E PROJETO DE DATA WAREHOUSE**. 4 ed. São Paulo. Editora Érica Ltda. 2009

SUPERSET. Disponível em <https://github.com/airbnb/superset>. Acesso em 5 nov 2016.

KNOWAGE. Disponível em < https://www.knowage-suite.com/site/home/> Acesso em 03 fev 2018.

ABDULLAH, A. **ANALYSIS OF MEALYBUG INCIDENCE ON THE COTTON CROP USING ADSS-OLAP (ONLINE ANALYTICAL PROCESSING) TOOL.** Computers and Electronics in Agriculture 69. p. 59 – 72, 2009.

RAI, A.; DUBKEY, V.; CHATURVEDI, K.K.; MALHOTRA, P.K.; **DESIGN AND DEVELOPMENT OF DATA MART FOR ANIMAL RESOURCES.** Computers and Electronics in Agriculture 64. p. 111 – 119. 2008