

DATA WAREHOUSE APLICADO NA GESTÃO DE DADOS AGRÍCOLA

Luiz Pedro Petroski (Estudante – Universidade Estadual de Ponta Grossa) E-mail: petroskilp@hotmail.com
Maria Salete Marcon Gomes Vaz (Professora Doutora – Universidade Estadual de Ponta Grossa) E-mail: salete@uepg.br

Resumo: Neste artigo os autores apresentaram um método para modelagem e geração de um esquema multidimensional lógico e físico. Também é apresentada duas definições para a elaboração de uma árvore de nós, e do esquema multidimensional lógico. A árvore de nós é a modelagem para a representação multidimensional que pode ser compreendida pelas partes interessadas (gestor agrícola e analista). Os passos seguintes correspondem a modelagem e regras de transformação do esquema multidimensional lógico e físico respectivamente. Através dos métodos abordados neste artigo notou-se que a partir dos metadados das bases de dados operacionais é possível estabelecer um padrão de modelagem que pode ser parcialmente automatizado. Também identificou a necessidade de adicionar atributos espaciais, para melhor detalhamento da dimensão espacial.

Palavras-chave: Data warehouse, dados agrícolas, modelagem.

DATA WAREHOUSE APPLIED IN AGRICULTURAL DATA MANAGEMENT

Abstract: In this paper the authors present a method for modeling and generating a logical and physical multidimensional scheme. It is also two definitions for an elaboration of a tree of nodes and a multidimensional schema. The tree of nodes is a modeling for a multidimensional representation that can be understood for the stakeholders (agricultural manager and analyst). The following steps correspond to a modeling and transformation rules of the logical and physical multidimensional scheme respectively. Through the methods discussed in this article, from a database of operational databases, it is possible to obtain a modeling model that can be partially automated. Also identified the need to add spatial attributes, to better detail the spatial dimension.

Keywords: Data warehouse, agricultural data, design.

1. INTRODUÇÃO

A variabilidade da produtividade das lavouras e com os fatores potencialmente limitantes a essa produtividade (VALENTINI e MOLIN, 2000). Assim áreas agrícolas são intensivamente monitoradas visando o entendimento da variabilidade espacial e do potencial de intervenções com as tecnologias disponíveis (MOLIN, 1999).

Diferentes Sistemas de Suporte a Decisão estão sendo desenvolvidos para facilitar a agro-industria. Estes Sistemas requerem metodos efetivos de armazenamento, como integração e métodos de análises. Data Warehouse é um tipo específico de banco de dados, sendo utilizado para integrar, acumular e analisar dados de diferentes fontes s (Calì et al. 2003). Usuários podem carregar informação de diferentes banco de dados em um data warehouse para análise combinada. De acordo com os requisitos, pode ser definido uma periodicidade para carregar os dados de uma base operacional para a base de análise (por exemplo diariamente, semanalmente, mensalmente).

A ferramenta responsável pelo processamento de consultas multidimensionais sobre um Data Warehouse é chamada Online Analytical Processing - OLAP. Esta produz indicadores estatísticos em diferentes níveis de detalhes. Os indicadores são calculados utilizando funções de agregação, como média, mínimo, máximo, permitindo ao usuário navegar através das hierarquias dimensionais, executando operações OLAP, como roll-up e drill-down.

2. DATA WAREHOUSE

Um Data Warehouse é composto tomando como base o conceito de fatos e dimensões. Fato é um ou mais indicadores mensuráveis, que podem ser agregados através de funções estatísticas (média, máximo, mínimo, soma, moda, desvio padrão). Um exemplo de um fato é a produtividade agrícola (valor numérico, mensurável). Uma dimensão é um eixo sobre o qual os fatos estão distribuídos, definindo o contexto do fato em questão. Por exemplo o tempo é uma dimensão sobre a qual os fatos estão distribuídos. Um Data Warehouse pode ter de 1 a n dimensões, sendo definidas na modelagem. Assim um valor de produtividade (fato) está contextualizado em um determinado espaço geográfico, tempo, cultura, máquina colhedora, operador (dimensões). Tanto fatos como dimensões, necessitam ser definidas a sua granularidade, ou seja, qual o nível de detalhamento de um fato, assim como as hierarquias de suas dimensões (KIMBALL e ROSS, 2013). A Figura 1 ilustra a representação de um cubo do fato produtividade e as dimensões cultura (soja, milho, feijão, trigo), talhão (talhão 1, talhão 2) e safra (safra 2014/2015, safra 2015/2016).

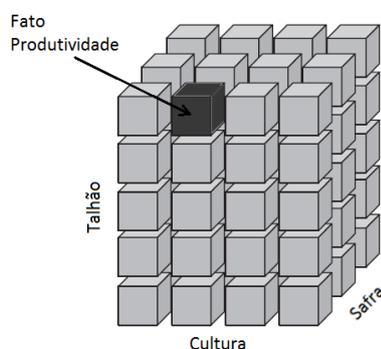


Figura 1 - Cubo de produtividade com as dimensões talhão, safra e cultura. Fonte: (PETROSKI e VAZ, 2016)

Os critérios de análise (cultura, talhão e safra) representam as dimensões do cubo. As dimensões podem ser estruturadas em hierarquias (Figura 2) e produz diferentes análises, pela combinação dos diferentes níveis das dimensões (PETROSKI e VAZ, 2016).

- Média de produtividade por talhão;
- Média de produtividade por talhão, por cultura e por safra;
- Média de produtividade por cidade, família da cultura e estação do ano.

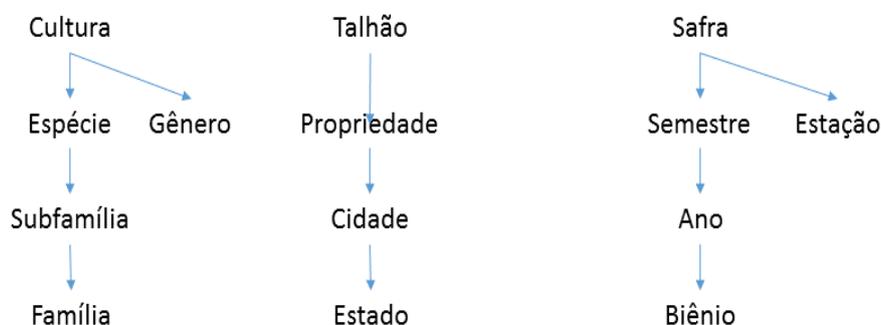


Figura 2 - Níveis de hierarquia das dimensões Fonte: (PETROSKI e VAZ, 2016).

3. MATERIAIS E MÉTODOS

Para contruir um cubo de dados para um Data Warehouse, deve-se iniciar pelo esquema multidimensional a partir das fontes de dados operacionais (Jensen et al., 2004; Romero e Abelló, 2008). O ponto principal da abordagem será explorar tanto a estrutura da fonte de dados (estrutura lógica), quanto as instâncias dos dados (os dados propriamente ditos) para determinar as dimensões e suas hierarquias, e os fatos.

Será utilizado algumas definições para a construção de uma árvore da estrutura da informação multidimensional. Esta árvore resultante será mapeada para o esquema multidimensional. O processo é baseado pela seleção de um nó e as medidas derivadas deste nó (indicadores), e os eixos de análise (dimensões).

Definição 1 (Mahboubi et al., 2013): Uma árvore O consiste em um conjunto de nós $E = \{e_1 \dots e_n\}$, onde n é o número de nós. Um nó e_i é definido como $e_i = \langle \text{nome}, \text{tipo} \rangle$, onde nome define o nome do nó, e o tipo define o tipo do nó ($\{\text{Complexo}, \text{Sequencial}, \text{Atributo}\}$). Um nó e_i do tipo atributo é um nó folha e definido como um tipo primitivo de dado.

Para a representação gráfica de uma árvore da definição 1 pode ser utilizada a tabela 1.

Tabela 1 - Representação gráfica da árvore de nós

Nó	Representação gráfica
Nó complexo,	●
Nó sequencial	[]
Nó atributo	○

Definição 2 (Mahboubi et al., 2013): Um esquema multidimensional M é representados por nós conceituais: fatos, medidas, níveis de dimensões e seus atributos correspondentes. É definido como $\langle e_{\text{fato}}, I, D, A \rangle$, onde e_{fato} define um nó fato, I é o conjunto de nós de medidas, D é o conjunto de níveis de dimensões a A é o conjunto de nós de atributos descritivos. O conjunto I é definido como $\{e_{\text{medida-0}}, \dots, e_{\text{medida-k}}\}$, onde k representa o numero de medidas. O conjunto D é definido como $\{e_{\text{dim-0,0}}, e_{\text{dim-i,j}}\}$, onde $e_{\text{dim-i,j}}$ representa o j -ésimo nível da i -ésima dimensão. O conjunto A é definido como $\{e_{\text{desc-}^0_0,0}, \dots, e_{\text{desc-}^z_{i,j}}\}$, onde $e_{\text{desc-}^z_{i,j}}$ representa o z -ésimo atributo descritivo do nível de dimensão $e_{\text{dim-i,j}}$.

Para a representação gráfica de um esquema multidimensional da definição 2 pode ser utilizada a tabela 2.

Tabela 2 - Representação gráfica de um esquema multidimensional

Nó	Representação gráfica
Fato	■
Atributo descritivo	■
Nível de dimensão	■
Medida	□

No próximo passo é a obtenção do esquema multidimensional físico a partir de um esquema multidimensional M . A implementação física segue os paradigmas relacional (KIMBALL e ROSS, 2013) e XML (Aouiche e Darmont, 2009; Mahboubi et al., 2009). O

esquema relacional pode ser obtido seguindo algumas regras de transformação (tabela 3) que auxiliam a obtenção do esquema multidimensional lógico e físico (declarações SQL).

Nível Lógico	Nível físico
Transformar todos níveis de dimensões de base em tabelas relacionais	Create table
Transformar cada nó de atributo descritivo em um atributo e adicionar a tabela de dimensão correspondente	Alter table add attribute
Transformar o nó fato em uma tabela relacional	Create table
Transformar os nós de medidas em um atributo numérico e adicionar a tabela de fatos	Alter table add attribute
Adicionar um atributo a tabela de fato para cada nó de nível de dimensão de base	Alter table add attribute
Adicionar uma chave primária na tabela de fatos	Alter table add constraint
Adicionar um chave primária para cada tabela de nível de dimensão	Alter table add constraint

Figura 3 - Regras de transformações SQL

3. RESULTADOS E DISCUSSÃO

Estudo de caso: Modelar um datawarehouse multidimensional para a agricultura de precisão no Brasil, possibilitando a análise temporal, de culturas e de localidades. Os dados estão organizados em uma base operacional CSV (comma-separated values) da seguinte maneira: produtividade (kg/ha), safra (ano), talhao (caractere), cultura(caractere). Outra base pode ser obtida, com informações do talhão: talhão (caractere), área (real). Por fim temos a base de dados composta pelos campos de fazenda (caractere), área (real) e região agrícola (caractere).

Utilizando a definição 1 para a modelagem, pode-se contruir a árvore de nós (figura 3) para a compor os primeiros passos para a modelagem multidimensional:

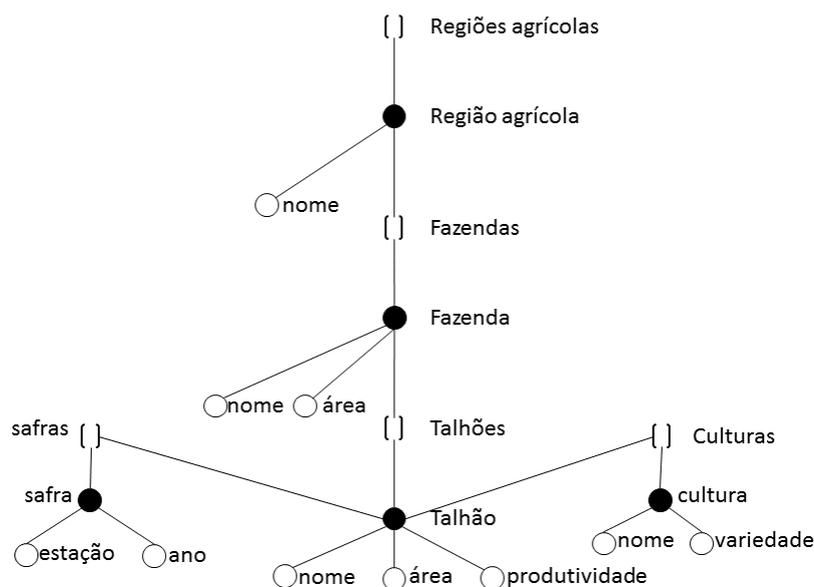


Figura 4 - Representação gráfica da árvore de nós para o estudo de caso proposto

O próximo passo, utilizando a definição 2, é construir o esquema multidimensional que para a modelagem do data warehouse (figura 4).

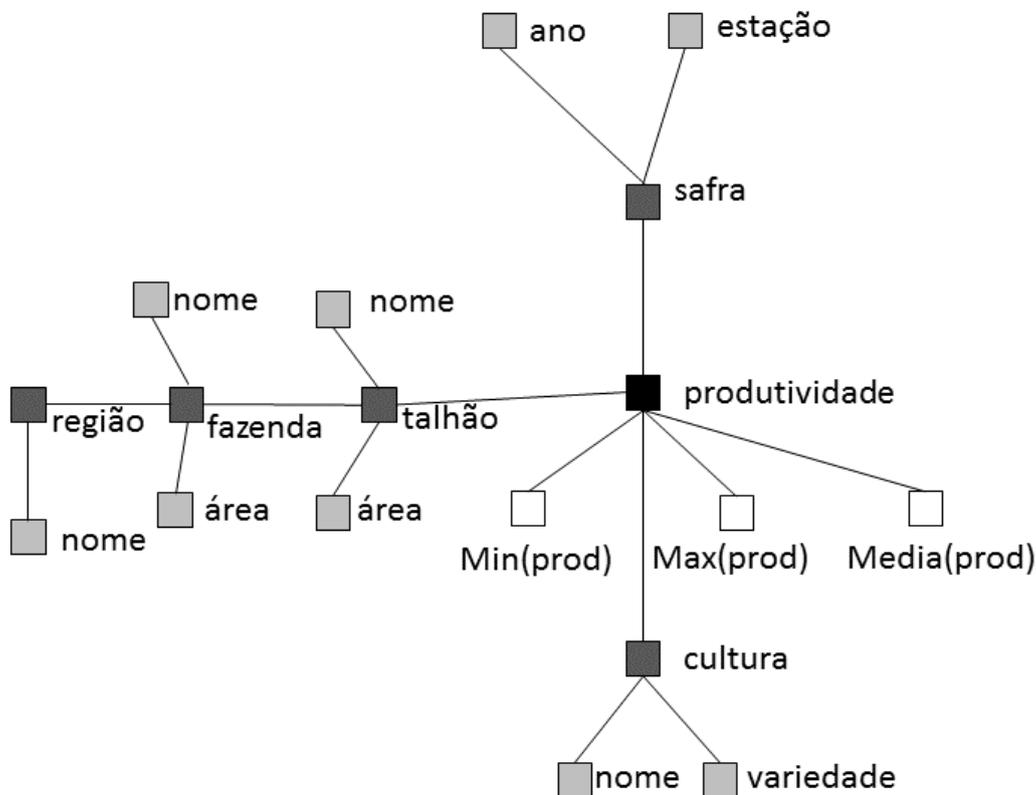


Figura 5 - Representação gráfica do esquema multidimensional do estudo de caso

E finalmente no último passo descrito na metodologia, definimos a modelagem multidimensional em nível físico, neste caso, representado pelo modelo da figura 5.

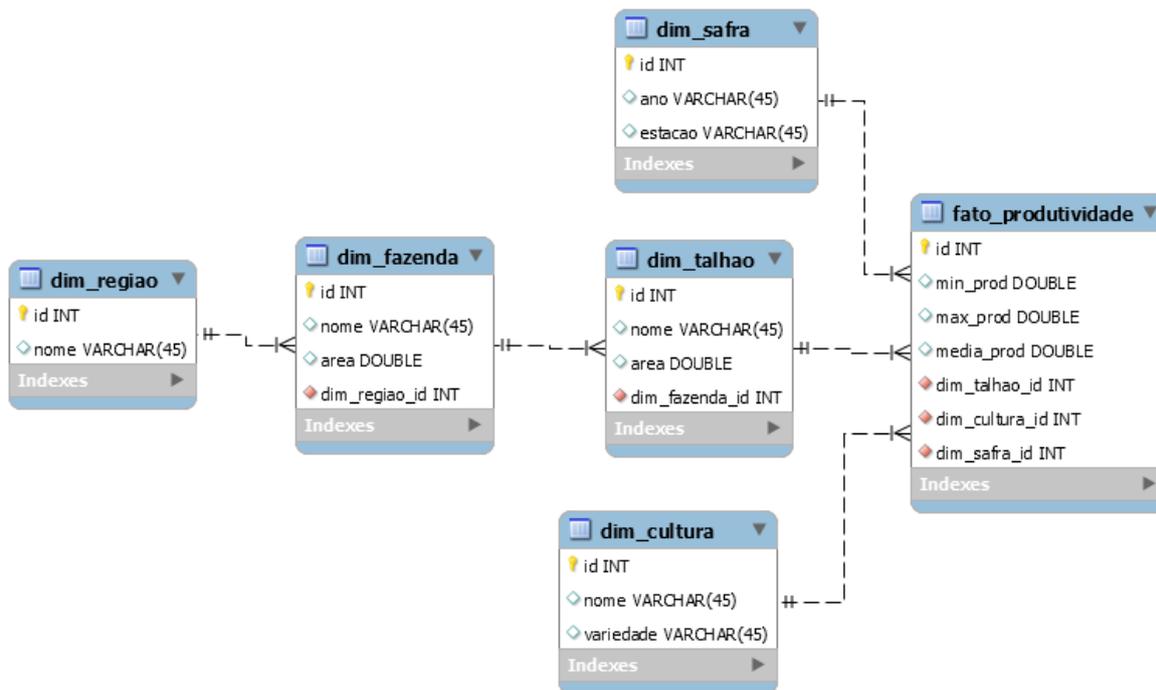


Figura 6 - Esquema multidimensional físico do estudo de caso

4. CONCLUSÕES

Neste artigo foi apresentado uma abordagem para a modelagem multidimensional de dados agrícolas para a análise de produtividade, nos contextos de safra, cultura e espaço.

O objetivo de uma modelagem multidimensional é construir um datawarehouse flexível, sendo possível a análise do gestor agrícola, nos mais diferentes cenários e entradas que uma ferramenta de suporte a decisão deve proporcionar.

Com a execução dos passos, notamos que é um processo que pode ser parcialmente automatizado, ou seja, através dos metadados de banco de dados operacionais, é possível extrair um padrão, afim de facilitar a modelagem de novos data warehouses agrícolas, assim como a construção de data marts (data warehouse específico para um domínio) para cada área do negócio agrícola. Por exemplo, pode ser criado um data mart para analisar as variedades mais produtivas do talhão 1 da fazenda nos últimos 10 anos.

Para trabalhos futuros propõe-se a adição de atributos espaciais para a descrição detalhada da dimensão espacial (ponto, linha, polígono), neste caso também se faz necessário a implementação de funções espaciais para as consultas multidimensionais, como por exemplo, junção, desjunção, perímetro, área, entre outras funções.

AGRADECIMENTOS

Os autores agradecem a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão da bolsa de estudo.

REFERÊNCIAS

Aouiche, K. e Darmont, J. Data mining-based materialized view and index selection on data warehouse. *Journal of Intelligent Information Systems*, 2009, 33(1), p. 65-93.

Calì A, Lembo D, Lenzerini M, Rosati R. Source integration for data warehousing. *Multidimensional databases*, Cemagref (2008) pp 361–392

Jensen, M. R., Holmgren, T., e Perderson, T. B. Discovering multidimensional structure in relational data. In Y. Kambayashi, M. Mohania, e W. Wob (Eds.), *Proceedings of the 6th International Conference on Data Warehousing and Knowledge Discovery*, Zaragoza, Spain, 2004 p. 138-148.

KIMBALL, R.; ROSS, M. *The Data Warehouse Toolkit*. 3. ed. [S.l.]: John Wiley & Sons, Inc., 2013.

PETROSKI, L. P.; VAZ, M. S. M. G. Modelagem multidimensional e espacial de um data Warehouse para base de dados agrícola In: *IV Conaitec: Congresso Agropecuário, Industrial e tecnológico do Paraná*, 2016, Ponta Grossa. Anais, 2016.

Mahboubi, H.; Bimonte, S.; Deffuant, G.; Chanet, J. P. & Pinet, F. Semi-automatic design of spatial data cubes from simulation model results *International Journal of Data Warehousing and Mining*, 2013

MOLIN, J.P. Geração de mapas de produtividade em projeto de avaliação da tecnologia de Agricultura de Precisão. IN: *Agrotec'99 – Conferência Internacional sobre Agropolos e*

Parques Tecnológicos Agroindustriais. Barretos, 1999, p. 231-235. Anais: <http://www.ibt-barretos.org.br>.

Romero, O., e Abelló, A. A survey of multidimensional modeling methodologies. *International Journal of Data Warehousing and Mining*, 2009, 5(2), p. 1-23.

VALENTINI, M.L., MOLIN, J.P. A Agricultura de Precisão nos Campos Gerais do Paraná. O Estado-da-arte em Agricultura de Precisão no Brasil. Piracicaba L. A. Balastreire, Coord. 2000. p. 88-92.