

Desenvolvimento de um sistema de gestão para apoio à tomada de decisão no agronegócio da região do Alto Paranaíba

Rodrigo Moreira¹, Dra. Adriana Zanella Martinhago¹,
Dr. Luis César Dias Drummond²

¹Instituto de Ciências Exatas e Tecnológicas – ICET
Universidade Federal de Viçosa *campus* Rio Paranaíba (UFV)
Caixa Postal 22 – 38.810-000 – Rio Paranaíba – MG – Brasil

²Instituto de Ciências Agrárias – IAP
Universidade Federal de Viçosa *campus* Rio Paranaíba (UFV).

moreira_r@outlook.com, adriana.martinhago@ufv.br, irriga@ufv.br

Abstract. *Organizations have computerized their processes through information systems, rural organizations have also followed this trend. With informatization, the amount of stored data has increased, and decision makers can support on information obtained from the data stored. The stored data is associated to the productivity of carrot and beet crops; and climate of the Rio Paranaíba region. These data are presented in forms of spreadsheets and are obtained from different sources, which makes the data analysis difficult. Therefore, the construction of a structured data repository is essential to enable the analysis, which aims to provide the information as a subsidy to decision makers in organizations.*

Resumo. *Organizações têm informatizado seus processos através de sistemas de informação, as organizações rurais também têm seguido essa tendência. Com a informatização, a quantidade de dados armazenados tem aumentado, e os tomadores de decisão podem apoiar em informações obtidas a partir dos dados armazenados. Os dados armazenados são associadas à produtividade das lavouras de cenoura e de beterraba; e clima da região de Rio Paranaíba. Estes dados são apresentados em formas de folhas de cálculo e são obtidos a partir de fontes diferentes, o que torna difícil a análise de dados. Portanto, a construção de um repositório de dados estruturado é essencial para permitir a análise, que tem como objetivo fornecer a informação como subsídio para os tomadores de decisão nas organizações.*

1. Introdução

Soluções de *Business Intelligence (BI)* ou Inteligência de Negócios, são cunhadas para padronização dos dados de forma adequada, em repositórios estruturados. A Inteligência de Negócios pode ser aplicada ao agronegócio, pois organizações desse setor tem investido em tecnologias de informação para diversos fins, como: monitorar o meio ambiente, automatização de processos, gerenciamento e etc. Cada novo sistema, fruto de investimentos em tecnologia da informação, são agentes criadores de dados, que posteriormente alimentarão ferramentas de *BI* [Barbieri 2001].

Criar um repositório de dados, *Data mart*, para armazenar dados relativos a produtividade de várias organizações do setor de agronegócio do Alto Paranaíba, juntamente com dados meteorológicos da região, proporcionará facilidade de acesso a pesquisadores, estudantes, professores e profissionais da área de agronomia às informações.

É importante possuir um repositório de dados irrepreensível, no que diz respeito a qualidade do seu conteúdo, portanto os dados devem ser tratados antes de compor do *Data mart*. Na fase de projeto do repositório de dados, é necessário atentar-se aos conceitos de modelagem multidimensional. Finalmente será possível apurar através de técnicas de *Business Intelligence*, níveis e a relação de produtividade contrastada com clima. A posse dessas informações viabiliza o processo decisório.

2. Data warehouse

Data warehouse (DW) é um banco de dados histórico, feito para o armazenamento dos dados extraídos de organizações. Para auxiliar os dados no apoio à tomada de decisão os dados devem ser armazenados em *DW*, eles são selecionados, integrados e organizados [Colaço 2004].

Segundo [Inmon et al. 1997] um *DW* deve possuir características como:

- não volatilidade: os dados em um *DW* só podem ser inseridos, atualizados e consultados;
- orientação à tópicos: *DW* armazena dados de um nicho específico, por exemplo: uma organização produtora de grãos, pode possuir um *DW*, onde este armazena, dados de clientes, informações sobre seus produtos e dados relativos ao clima;
- integrado: um *DW* deve ser capaz de armazenar dados que são oriundos de diversos sistemas transacionais da organização. Entretanto, é possível que cada sistema departamental utilize uma convenção para armazenar em seu banco de dados relacional esses dados operados diariamente. Quando esses dados são armazenados em *DW* pode haver incompatibilidades. Tome como exemplo o campo Estado (Unidade da Federação), em algum sistema pode-se utilizar a convenção de siglas, por exemplo: SP, ao passo que no *DW* pode ser usado um campo texto para o atributo em questão, isto é: São Paulo. O *DW* deve ser capaz de armazenar os dados fazendo uso de uma única convenção;
- varia com o passar do tempo: no *DW* os dados passam por atualizações no decorrer do tempo. Tome como exemplo um *DW* que continha dados de clientes solteiros no ano de 2009. Alguns anos depois, se parte ou todos esses clientes estiverem casados, o *DW* deve ser capaz de oferecer os dados para as duas circunstâncias;

2.1. Data mart

Data mart é um conjunto de dados flexíveis, baseados na maioria das vezes como dados atômicos, estes podem ser extraídos através de uma fonte operacional. São apresentados como modelo simétrico, resistente em ocasiões que o usuário faz consultas inesperadas. É uma fatia do *DW*, dedicado a manter dados de um único processo do negócio [Kimball and Merz 2000].

A Figura 1 ilustra um exemplo de um *Data mart*, tem-se Vendas, Clientes e Compras, cada um representa um processo de negócio específico [Laudon and Laudon 1999]. *Data marts* são específicos para um departamento da organização, portanto é uma

especialização do *DW*, feito para o armazenamento de dados em uma escala menor [Barbieri 2001].

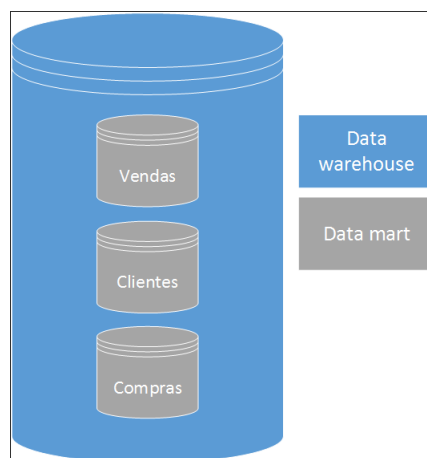


Figura 1. Abstração de um *Data warehouse*.

3. Extração, Transformação e Carga (ETC)

Segundo [Kimball and Merz 2000] pode-se determinar os objetivos do ETC como: (1) entregar dados de forma mais eficaz para as ferramentas de usuário final; (2) adicionar valor aos dados nas etapas de limpeza e padronização; (3) proteger e documentar a linhagem dos dados.

Extração, pode resumidamente ser conceituada como etapa de extração de dados provenientes dos sistemas transacionais. A etapa de transformação é responsável pela tarefa de padronização dos dados, para que a base de dados se torne inequívoca e completa. Carga, nessa etapa concentram-se atividades para o carregamento dos dados já tratados na estrutura multidimensional do armazém de dados [El-Sappagh et al. 2011].

4. Trabalhos Relacionados

- **Uso de Ferramentas de *Business Intelligence* na Análise de Desempenho de uma Empresa de Agronegócios**

[Lima and Boscaroli 2012] propõe a descrição de uma ferramenta para análise de desempenho em uma empresa, com o uso do processo de *BI*. Focada no agronegócio a empresa desenvolve atividades como comercialização de insumos e beneficiamento de grãos.

Para realização do trabalho foi investigada características da empresa para facilitar o entendimento dos processos internos e externos. Elaborou-se métricas de desempenho que são essenciais para análise corporativa do agronegócio. Fora arquitetado um *Data mart*, com foco na análise dos indicadores de desempenho. Também elaborou-se o agente ETC, para refinar a base de dados com o que realmente é importante e contribui positivamente com o processo de *BI*.

O trabalho descrito é relacionado com o desenvolvido na questão de criação de *Data mart*, para ser usado pela ferramenta de *BI*, para obtenção de informações que venham ser úteis na tomada de decisão. Possui divergência com o proposto, pois no trabalho

relacionado é criado um *Data mart* para armazenamento de dados de venda, no desenvolvido foi utilizado um *Data mart* para armazenar dados de clima e produção.

- **Representação de comercialização agropecuária através de modelo de *Data warehouse***

O trabalho de [Correa 2009] apresenta um estudo sobre modelos de dados dimensionais aplicados à definição de *DW* e *Data marts* com foco em agronegócio. É abordado questões sobre análise de informações e criação de *Data marts* para armazenar dados de pecuária e grãos. Para viabilizar a realização de processamento analítico online, utilizou-se como cadeias de comercialização pecuária e grãos como soja e milho. Justifica-se o uso dessas cadeias a importância no agronegócio brasileiro.

Com o desenvolvimento do trabalho foi possível fazer análises de mercado. Pôde ser feito cruzamento de informações, avaliar os impactos das variações de preços. O que apoia a tomada de decisões. O desenvolvimento de um *Data warehouse* possibilitou o armazenamento centralizado de dados sobre comercialização e produção.

Com relação ao trabalho desenvolvido, esse se relaciona no quesito de modelagem de armazém de dados históricos com foco no agronegócio. E no uso do Sistema de Gerenciamento Banco de Dados (SGBD) para gerenciamento dos dados armazenados.

5. Métodos

Baseado na Figura 2 é apresentado os desdobramentos das etapas que compõe o ciclo de desenvolvimento do trabalho proposto.

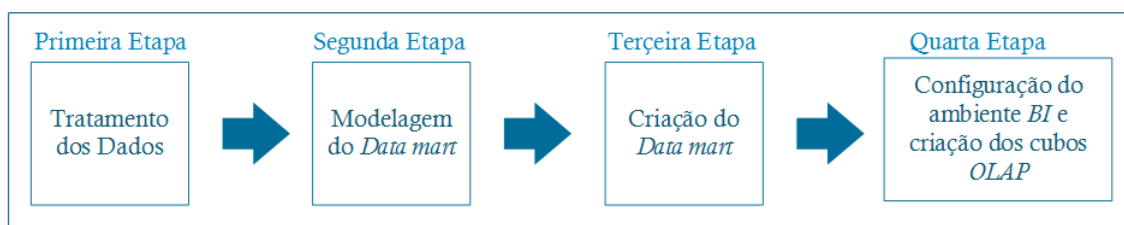


Figura 2. Etapas do desenvolvimento do Trabalho.

A primeira etapa representada na Figura 2 contempla a atividade de tratamento dos dados. Eles foram obtidos da Estação Meteorológica instalada no *Campus I* UFV Rio Paranaíba e foram exportados em forma de planilhas eletrônicas. O sistema da Estação Meteorológica é programado para coletar dados a cada 60 min, esses procedimentos são realizados na frequência mencionada desde janeiro de 2009. Os dados contidos nas planilhas eletrônicas estão dispostos em forma de linhas, onde cada linha representa um registro. Cada registro contém: Data, Hora, Temperatura Máxima, Temperatura Mínima, Velocidade do Vento, Direção do Vento, Humidade Relativa do Ar, Evapotranspiração. Também foram tratados dados quantitativos da produtividade de culturas de cenoura e beterraba de fazendas da zona rural dos municípios de Rio Paranaíba – MG e São Gotardo – MG, cedidos pelo Grupo *Sekita* Agronegócios.

A etapa de tratamento dos dados consiste em retirar campos nulos causados por eventuais falhas técnicas ou não preenchimento, no caso das tabelas de produtividade. A relevância do tratamento é justificada pela necessidade de armazenar dados íntegros e

confiáveis no *Data mart*. Para o tratamento foi utilizada a ferramenta *Data Integration* da suíte *Pentaho*. Pois permite-se escolher registros e campos da planilha eletrônica a serem carregados pelo *Data mart*, também ignorar campos nulos e aqueles que não compõem o *Data mart*, seja na tabela de Fatos ou em tabelas Dimensão.

A segunda etapa contém a atividade de modelar o *Data mart*. Para modelagem foi utilizado a ferramenta *open source MySQL Workbench*. Na modelagem foi adotado a abordagem *Star Schema*, para que o banco de dados seja consolidado na forma multidimensional como propõe [Colaço 2004].

Os benefícios de um *Star Schema* ultrapassa os de um modelo normalizado, pelo fato das consultas no armazém de dados ser apenas de leitura [Teorey et al. 2014]. Em adição, [Weininger 2002] propõe técnicas para melhorar as consultas em modelos *Star Schema*, que podem ser exploradas por duas vertentes; (1) *hash push down*, (2) *bit-vector push down*. Ambas objetivam a redução de linhas, ao realizar operação do tipo *Join*.

A terceira etapa consiste em criar e povoar o *Data mart* com base na modelagem desenvolvida. Foi utilizado o Sistema Gerenciador de Banco de Dados (SGBD) *PostgreSQL*, pelo seu bom desempenho, por ser *open source* e apresentar boa compatibilidade com a suíte *Pentaho*, principalmente no que diz respeito a *plugins*¹ de conexão com os módulos do *Pentaho* [Pires et al. 2006].

Após consolidado o *Data mart* e com os dados carregados, segue as atividades da quarta etapa. Nesta etapa foi configurado o ambiente *Pentaho Business Analytics* para possibilitar análises. Também foi criado cubos *OLAP*² como parte da solução de *BI*, com suas características relacionadas a granularidade e métricas. Para criação dos cubos foi utilizado o módulo *Mondrian* da suíte *Pentaho*. Após criado cubos *OLAP*, eles puderam ser publicados para o *Pentaho BI Server* para que fosse possível visualizá-los e analisá-los, seja de forma gráfica ou tabular.

6. Resultados e Discussões

Após seguir sistematicamente a metodologia, foi desenvolvido o modelo lógico do *Data mart* que pode ser visualizado na Figura 3. Para tal, foi utilizado a ferramenta *MySQL Workbench*. No modelo, pode ser visualizado a tabela de fatos; “tableFato” como centro do esquema e, as demais tabelas: “tableData”, “tableClima”, “tableGleba”, “tableCultura” e “tableFazenda”. Todos atributos da tabela foram criados para armazenar valores de campos provenientes das planilhas de clima e produção. A tabela de fatos possui a chave estrangeira (*Foreign Key*) de cada dimensão que a habilita contatar qualquer tabela na borda do esquema, possui também um atributo de medida, conceitualmente denominado como métrica; no caso do trabalho desenvolvido é a produtividade. As demais tabelas, são tabelas de dimensão, elas possuem seus atributos e uma chave primária (*Primary Key*) para poder ser ligada a tabela de fatos, dessa forma é possível apurar a produtividade por alguma dimensão.

¹Também conhecido como módulo de extensão, utilizado para adicionar função a outros programas maiores.

²*On-line Analytical Processing (OLAP)* ou Processamento Analítico On-Line é conjunto de técnicas usadas para tratar informações contidas em repositório de dados estruturados *DW* ou *Data mart* [Colaço 2004]

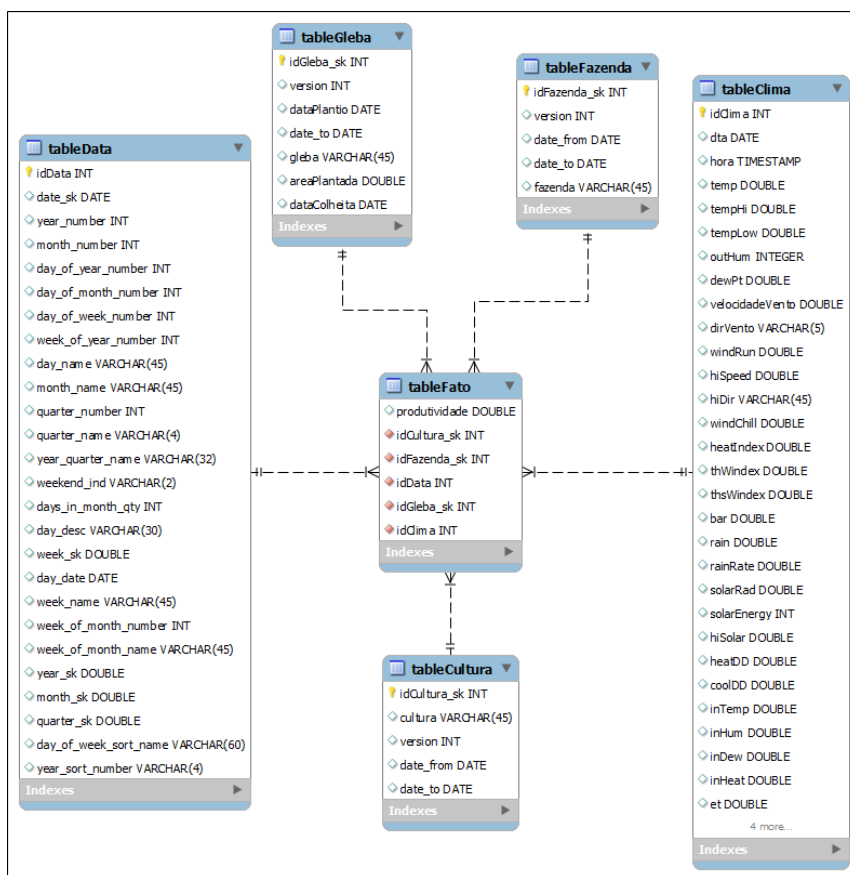


Figura 3. Modelo lógico do Data mart.

A Figura 4 ilustra o processo de Extração, Transformação e Carga (ETC) modelado para armazenar os dados na tabela de fatos do Data mart. O primeiro componente foi configurado para realizar a leitura linha a linha da planilha de produção. Na medida que são lidas elas seguem para o componente “FiltroLinhasNulas” que trata a eventualidade de campos nulos, se houver a linha é descartada pelo componente “DescartaLinhas”. Quando a linha lida é íntegra, ela segue para o componente “DimensionamentoCultura”, esse componente foi configurado para comparar o nome da cultura da planilha com o nome da cultura na “tableCultura”, quando for igual esse componente encaminha o “id-Cultura” para o próximo componente.

O componente “DimensionamentoGleba” foi configurado para comparar o nome da Gleba registrado na planilha com o nome da Gleba armazenado na “tableGleba”, quando for igual esse componente encaminha o “idGleba” para o próximo componente. O componente “DimensionamentoFazenda” segue a mesma sistemática, compara o nome da Fazenda registrado na planilha com o nome da Fazenda armazenado na “tableFazenda”, quando for igual esse componente encaminha o “idFazenda” para o próximo.

O componente “DimensionamentoData” foi configurado para vincular uma data da “tableData” com a data de plantio da cultura. O componente foi configurado para comparar se a data da tabela é igual a alguma data de plantio registrada na planilha de produção, quando houver, o componente encaminha para o próximo o “idData”. No próximo componente; “DimensionamentoClima”, foi configurado para comparar se a data

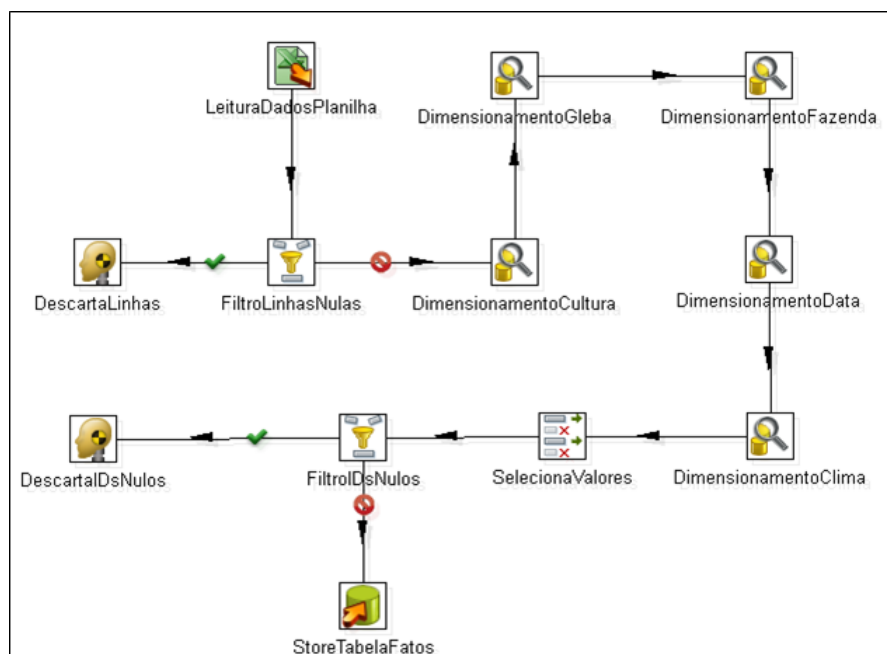


Figura 4. Transformação da tabela Fato.

do clima contida na tabela “tableClima” é igual a data de plantio registrada na planilha de produção, quando forem iguais, o componente encaminha o “idClima”.

Foi criado um componente “SelecionaValores”, esse componente é responsável por selecionar os campos que serão armazenados na tabela de fatos. O componente seleciona: “produtividade”, “idFazenda”, “idClima”, “idGleba”, “idData” e “idCultura” para poder compor a tabela de fatos do *Data mart*. Após selecionado os campos, o componente “FiltroIDsNulos” verifica a ocorrência de “ids” nulos. A ocorrência de algum, faz com que o registro dimensionado seja descartado, para assegurar as restrições de integridade de chave no *Data mart*.

Na ocorrência de “ids” íntegros encaminha-se o fluxo de dados para o último componente, “StoreTabelaFatos”. Nesse componente foi elaborado consulta *SQL* para fazer a inserção na “tableFato”. O registro armazenado na tabela de fato consegue através das chaves estrangeiras (*Foreign Key*) vincular uma produtividade; métrica escolhida com: Fazenda, Gleba, Cultura, Data e Clima. Com essa estrutura do *Data mart* pode-se aplicar técnicas de consultas *OLAP* através de ferramentas *BI*.

Com o desenvolvimento do sistema é possível apurar pela ferramenta *Pentaho Business Analytics* a produtividade das culturas de cenoura e beterraba da região do Alto Paranaíba de forma tabular ou gráfica. De maneira semelhante é possível visualizar uma condição climática e a produtividade de uma determinada época. Esse tipo de informação é essencial para os gestores de organizações rurais, pois os possibilitam avaliar investimentos ao plantar determinada cultura em uma época específica. As informações que um sistema de *BI* proporciona aos tomadores de decisão são subsídios somadores para uma decisão acertada.

As ilustrações subsequentes são cópias autênticas dos gráficos gerados pelo *Pentaho Business Analytics*. A Figura 5 ilustra o nome das fazendas e o total de produção

da cultura de cenoura. A fazenda “Lote46” apresenta maior produção dessa cultura. A Figura 6 ilustra de forma gráfica a produtividade de cenoura e a condição climática que a cultura foi submetida desde o plantio até a colheita. A Figura 7 ilustra o gráfico da produtividade em caixas por hectare da cultura de cenoura. Os detalhes da barra contidos na caixa preta retrata que, em 15.8 hectare de área plantada a produção total foi de 6.172 caixas por hectare.

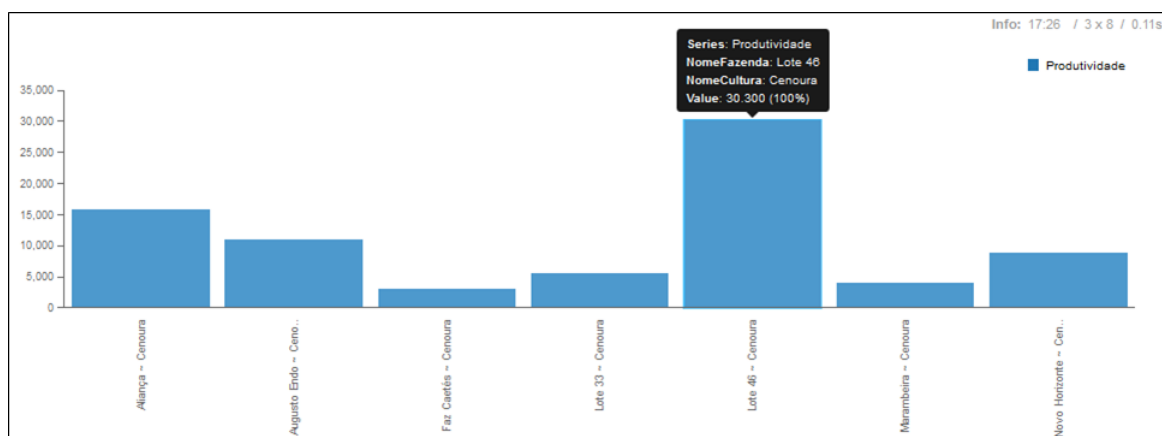


Figura 5. Produtividade: Fazenda × Cultura

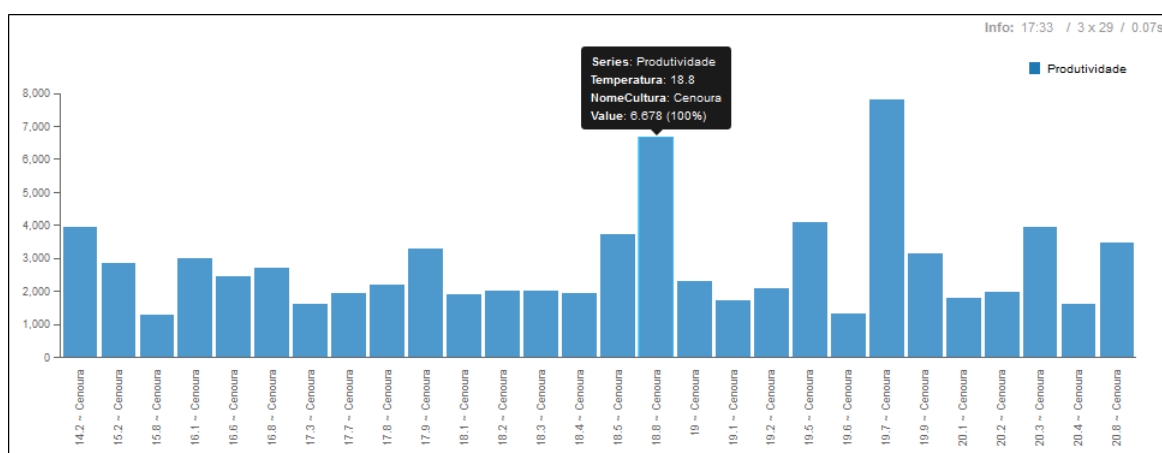


Figura 6. Produtividade: Cultura × Temperatura

O gráfico ilustrado na Figura 8 mostra a distribuição média do clima da região do Alto Paranaíba dos anos de 2008 à 2011. Os detalhes de um ponto específico podem ser vistos na caixa preta, onde é retratado que na data 23-04-2010 a temperatura média foi de 23,605 °C.

7. Conclusões

Foi possível analisar a produtividade que é dada em caixas por hectare por diversas dimensões. A possibilidade de criar gráficos com o arrastar de componentes mostrou a potencialidade da análise multidimensional de dados para apoio à decisão. A geração de gráficos de clima possibilita adquirir informações sobre vários aspectos que compõe uma

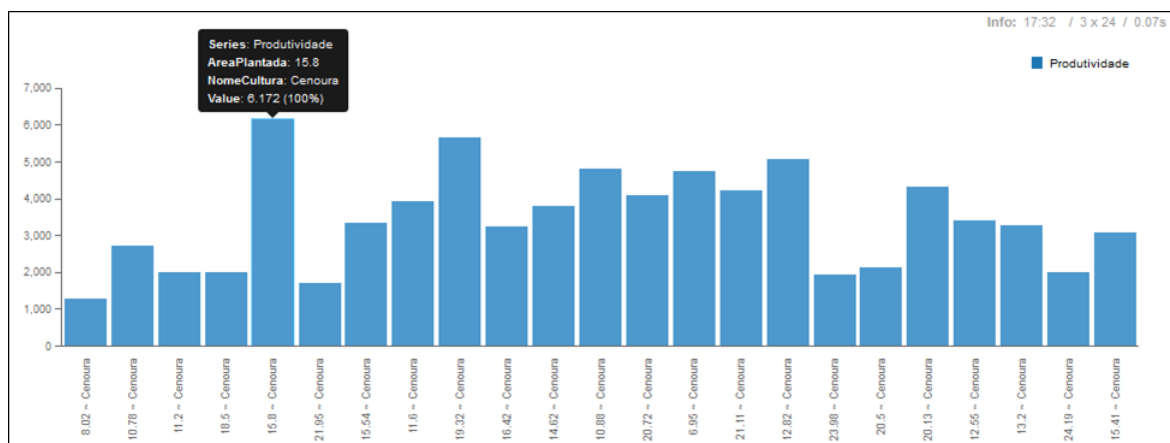


Figura 7. Gráfico Produtividade Cultura × Área Plantada.

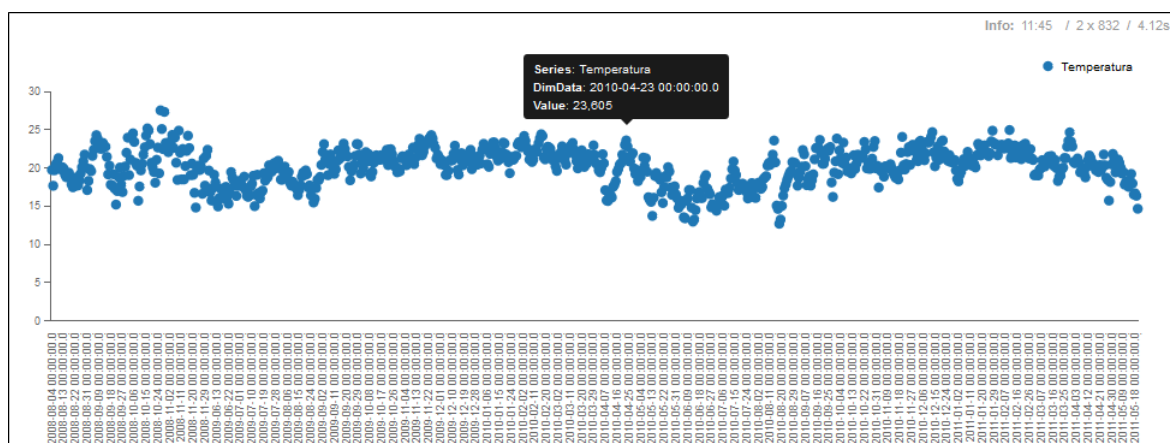


Figura 8. Gráfico Temperatura média dos anos.

condição climática. Não somente temperatura média, mas outras informações que profissionais da área de agronomia possuem afinidade como: radiação, índice de umidade, ponto de orvalho e etc. A visualização dessas e demais características é possível. Informações desse nível é subsídio somador no momento da decisão para saber o quanto investir na plantação de determinada cultura. Essas informações também podem contribuir com pesquisas na região do Alto Paranaíba.

No que tange o desenvolvimento do trabalho, foi possível construir o modelo lógico do *Data mart* que mostrou-se eficiente e funcional para análises propostas. A criação do *Data mart* foi concluída após utilizar Linguagem de Definição de Dados (LDD) *SQL* em um banco de dados *PostgreSQL*. Os módulos de ETC realizaram o devido tratamento e carregamento dos dados oriundos das planilhas eletrônicas. É possível assegurar que não havia dados incoerentes no *Data mart*. Com a utilização do *Mondrian* foi possível criar e publicar cubos *OLAP* utilizados no ambiente de análise. Para análise dos dados, foi configurado o ambiente *BI* (*Pentaho Business Analytics*) que possibilita diagnóstico dos dados de forma gráfica e tabular através do *Browser*. Em suma, o montante das etapas consolidam a criação de um sistema que consegue apoiar à tomada de decisão de gestores em empresas rurais da região do Alto Paranaíba.

Referências

- Barbieri, C. (2001). *BI-business intelligence: modelagem e tecnologia*. Axcel Books.
- Colaço, M. (2004). *Projetando Sistemas de Apoio à Decisão Baseados Em Data Warehouse*. AXCEL BOOKS, 1st edition.
- Correa, F. E. (2009). Representação de comercialização agropecuária através de modelo de data warehouse.
- El-Sappagh, S. H. A., Hendawi, A. M. A., and Bastawissy, A. H. E. (2011). A proposed model for data warehouse {ETL} processes. *Journal of King Saud University - Computer and Information Sciences*, 23(2):91 – 104.
- Inmon, W. H., Zachman, J. A., and Geiger, J. G. (1997). *Data Stores, Data Warehousing and the Zachman Framework: Managing Enterprise Knowledge*. McGraw-Hill, Inc., New York, NY, USA, 1st edition.
- Kimball, R. and Merz, R. (2000). *Data webhouse: construindo o data warehouse para a Web*. Campus.
- Laudon, K. C. and Laudon, J. P. (1999). *Sistemas de informação: com Internet*. LTC Editora.
- Lima, V. M. d. and Boscarioli, C. (2012). Uso de ferramentas de business intelligence na análise de desempenho de uma empresa de agronegócios. pages 420–431.
- Pires, C. E. S., Nascimento, R. O., and Salgado, A. C. (2006). Comparativo de desempenho entre banco de dados de código aberto. *Escola Regional de Banco de Dados, Anais da ERBD 06*.
- Teorey, T., Lightstone, S., Nadeau, T., and Jagadish, H. V. (2014). *Projeto e Modelagem de Banco de Dados*. Rio de Janeiro, RJ, Brazil, 2nd edition.
- Weininger, A. (2002). Efficient execution of joins in a star schema. In *Proceedings of the 2002 ACM SIGMOD International Conference on Management of Data*, SIGMOD '02, pages 542–545, New York, NY, USA. ACM.