

UNIVERSIDADE ESTADUAL DE PONTA GROSSA
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO APLICADA

LUIZ PEDRO PETROSKI

**Uma arquitetura para integração de ambientes data warehouse, espacial e
agricultura de precisão**

Ponta Grossa

2017

**UNIVERSIDADE ESTADUAL DE PONTA GROSSA
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO APLICADA**

LUIZ PEDRO PETROSKI

**Uma arquitetura para integração de ambientes data warehouse, espacial e
agricultura de precisão**

Dissertação apresentada ao Programa de Pós-Graduação em Computação Aplicada, curso de Mestrado em Computação Aplicada da Universidade Estadual de Ponta Grossa, como requisito parcial para obtenção do título de Mestre.

Orientador: Prof^a Dr^a. Maria Salete M. G. Vaz

Ponta Grossa

2017

Ficha Catalográfica
Elaborada pelo Setor de Tratamento da Informação BICEN/UEPG

P497 Petroski, Luiz Pedro
 Uma arquitetura para integração de
 ambientes data warehouse, espacial e
 agricultura de precisão/ Luiz Pedro
 Petroski. Ponta Grossa, 2017.
 79f.

 Dissertação (Mestrado em Computação
 Aplicada - Área de Concentração:
 Computação para Tecnologias em
 Agricultura), Universidade Estadual de
 Ponta Grossa.
 Orientadora: Prof^a Dr^a Maria Salete M.
 G. Vaz.

 1.Data Warehouse. 2.OLAP. 3.SIG.
 4.Agricultura. 5.Sistemas de Suporte a
 Decisão. I.Vaz, Maria Salete M. G.. II.
 Universidade Estadual de Ponta Grossa.
 Mestrado em Computação Aplicada. III. T.

CDD: 631

TERMO DE APROVAÇÃO

Luiz Pedro Petroski

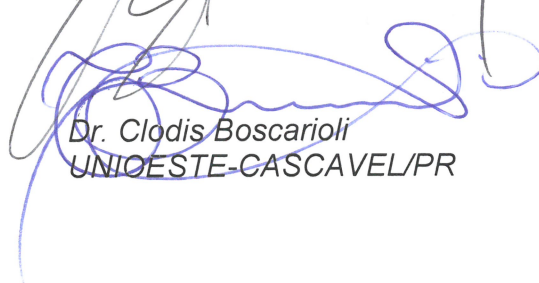
**“UMA ARQUITETURA PARA INTEGRAÇÃO DE AMBIENTES
DATA WAREHOUSE, ESPACIAL E AGRICULTURA DE PRECISÃO”.**

Dissertação aprovada como requisito parcial para obtenção do grau de Mestre no Programa de Pós-Graduação em Computação Aplicada da Universidade Estadual de Ponta Grossa, pela seguinte banca examinadora:

Orientadora:


Dr.^a Maria Salete Marcon Gomes Vaz
UEPG


Dr. Marcelo Giovanetti Canteri
UEL


Dr. Clodis Boscarioli
UNIOESTE-CASCAVEL/PR

Ponta Grossa, 14 de março de 2017.

Dedico esta dissertação à minha Mãe Eneusira (*in memorian*), que dedicou a sua vida aos seus filhos, com muito amor, carinho e dedicação. Também dedico ao meu Pai Luiz, um herói e maior exemplo, a quem cresci admirando e me espelhando em seu caráter, humildade, honestidade e retidão.

Dedico à minha noiva e futura esposa Pâmella, pelo apoio incondicional e constante incentivo, ao meu irmão Vitor e minha orientadora Maria Salete.

O apoio de todos foi determinante para a realização deste trabalho. A eles, meu muito, muito obrigado.

Agradecimentos

A DEUS, por iluminar o meu caminho e me dar forças para seguir sempre em frente.

Aos meus pais, pois desde pequeno me ensinaram o valor do conhecimento, me proporcionando as oportunidades que me levaram a chegar até aqui. Ao meu pai Luiz, quem sempre admirei, e até hoje admiro, que nunca mediu esforços para dar a seus filhos uma vida digna, e a oportunidade de almejar novos horizontes, os quais vislumbro neste momento. De modo especial à minha Mãe Eneusira (*in memorian*), o meu maior motivo de estar aqui, não exclusivamente por ser minha genitora, e sim por ser "Mãe" em todos os sentidos que essa palavra possa expressar. Em momentos que nem eu mesmo acreditava em mim, a única pessoa a acreditar fiel e cegamente era você Mãe. Agradeço meus pais por serem o meu referencial, minha "casa".

À minha noiva e futura esposa Pâmella, pelo apoio dispensado durante toda essa jornada, pelas horas que abdiquei de estarmos juntos, afim de atingir meu objetivo, sendo compreensível e dando suporte em todos os momentos.

Ao meu irmão Vitor, que escolheu a mesma área para seguir, e que me "aturou" durante todo esse tempo de convivência, e por compreender os momentos em que estive ausente.

Ao meu colega, sócio e amigo Albino Szesz Junior, por todas as dificuldades, parcerias e projetos que juntos realizamos durante todo esse processo.

À minha orientadora Prof^a Dr^a Maria Salete Marcon Gomes Vaz, pelo incentivo e confiança que depositou em mim, desde o primeiro momento em que nos conhecemos. Nunca esqueço, o dia que fui em sua sala para pedir isenção da taxa de inscrição para participar de um evento. A partir deste momento, foram muitos projetos, extensão, iniciação científica, oportunidades acadêmicas e profissionais, os quais serei eternamente grato.

Aos meus colegas de turma, os quais dividimos momentos difíceis "abraçados na madrugada" com trabalhos em grupo, dúvidas e prazos. Mas também momentos alegres, de descontração, e da incrível sensação de dever cumprido.

Ao professor Luiz Cláudio Garcia pelo auxílio em promover o contato, que foi indispensável para a fomentar este trabalho.

Ao Engenheiro Agrônomo Allison Fornari, pelo pronto atendimento, e o fornecimento de dados agrícolas, possibilitando a continuidade deste trabalho.

Ao Programa de Pós Graduação em Computação Aplicada, que conduz com muita eficácia e seriedade as suas atividades, proporcionando um ambiente de ensino de qualidade e muitas oportunidades de pesquisa.

Agradeço a todos os professores por me proporcionar o conhecimento não apenas racional, mas a manifestação do caráter e afetividade da educação no processo de formação profissional, por tanto que se dedicam a mim, não somente por terem me ensinado, mas por terem me feito aprender. A palavra mestre, nunca fará justiça aos professores dedicados aos quais sem nominar terão os meus eternos agradecimentos.

À Universidade Estadual de Ponta Grossa, pela oportunidade e estrutura cedida para os estudos e a Capes pela bolsa de estudos.

A todos que direta ou indiretamente fizeram parte da minha formação, o meu muito obrigado.

"A menos que modifiquemos a nossa
maneira de pensar, não seremos capazes
de resolver os problemas causados pela
forma como nos acostumamos a ver o
mundo"

Albert Einstein

O objetivo desta dissertação é apresentar uma proposta de integração entre agricultura de precisão, *Data Warehouse*/OLAP e SIG. A integração deve utilizar componentes abertos e extensíveis, modelagem agrícola para suporte a decisão, suporte a dados geográficos, interface de comunicação entre os componentes e a extensão de soluções existentes de SIG e *Data Warehouse*. Como resultado da integração foi definido uma arquitetura aberta e extensível, integrada, com uma modelagem de *Data Warehouse* agrícola espacial, que permite o suporte a tomada de decisão para o planejamento e gestão do manejo das práticas da agricultura de precisão. Desta forma as tecnologias e ferramentas utilizadas são abertas e permitem a implementação e extensão de suas funcionalidades para adequar ao cenário agrícola de tomada de decisão. Para realizar a integração foi utilizado os dados oriundos de uma fazenda localizada em Piraí do Sul/PR, a qual utiliza um software proprietário para o gerenciamento de dados. Os dados foram exportados para o formato SHAPEFILE, e através do processo realizado pela ferramenta de ETL, foram extraídos, transformados e carregados para a base de dados analítica. Também como fonte de dados sobre as fronteiras políticas das regiões rurais do Brasil, foi utilizado dados do IBGE. A base de dados analítica foi modelada e implementada em um SGBD PostgreSQL com a extensão PostGIS para suportar os dados geográficos. Para prover o serviço de consultas OLAP, foi utilizado o servidor Geomondrian. A aplicação foi estendida do projeto Geonode, onde foi implementado as funcionalidades analíticas, e a interface entre a aplicação e o servidor OLAP, foi realizada pela API Mandoline e a biblioteca OLAP4J. E por fim a interface foi implementada por meio de bibliotecas javascript para a criação de gráficos, tabelas e mapas. Como principal resultado, obteve-se uma arquitetura para integração de datawarehouse, operações OLAP, dados espaciais e agricultura, bem como definição do processo de ETL e a interface com o usuário.

Palavras-chave: Data Warehouse, OLAP, SIG, Agricultura, Sistemas de Suporte a Decisão, modelagem, banco de dados

Abstract

The aim of this work is to present a proposal of integration between Precision Agriculture, Data Warehouse / OLAP and GIS. The integration should use extensible and open components, agricultural modeling for decision support, geographical data support, communication interface between components, extension of existing GIS and Data Warehouse solutions. As a result of the integration, an open and extensible architecture was defined, with a spatial agricultural data warehouse modeling. In this way the technologies and tools are open and allow the implementation and extension of its functionalities to adapt to the agricultural scenario of decision. In order to perform the integration, the data were obtained from a farm in the city of Pirai do Sul/PR, which uses proprietary software for data management. Data was exported to the SHAPEFILE format, and through the process performed by the ETL tool, was extracted, transformed and loaded into the analytical database. Also as a source of political boundaries data of rural regions of Brazil, data from the IBGE were used. The database was modeled and implemented by PostgreSQL DBMS with the extension PostGIS to support spatial data. To provide the OLAP query service, was used the Geomondrian server. The application was extended from the Geonode project, where it was implemented Analytic functionalities, and the interface between the application and the OLAP was performed by the Mandoline API and the OLAP4J library. And finally the interface was implemented through javascript libraries for creating charts, tables and maps. As principal result, an architecture was obtained for Data Warehouse integration, OLAP operations, agricultural and spatial data, as well as ETL process definition and the user interface.

Keywords: Data Warehouse, OLAP, GIS, Agriculture, Decision Support Systems, modeling, database

Lista de ilustrações

1	Fluxo da informação do ciclo produtivo agrícola, aplicando processos de análise e manejo na agricultura de precisão.	23
2	Arquitetura DW	24
3	Modelo estrela e cubo	25
4	Esquema de obtenção das medidas de produtividade do processo produtivo através dos sensores da colhedora	25
5	Operações OLAP	28
6	Sobreposição de camadas	29
7	Esquema de representação espacial da estrutura matricial	30
8	Esquema de Representação Espacial da Estrutura Vetorial	30
9	Diagrama de classe - especialização GM_Object	32
10	Modelo de fluxo de trabalho geral para a criação de um Data Warehouse espacial	36
11	Diagrama de implementação da camada Aplicação WEB	39
12	Solução para realização dos filtros e agregações via consulta	41
13	Solução para realização dos filtros e agregações via servidor	41
14	Método proposto baseado em tecnologias abertas e extensíveis para integração de um ambiente <i>Data Warehouse</i> , <i>OLAP</i> e SIG para agricultura	52
15	Arquitetura proposta para a integração de <i>Data Warehouse</i> e SIG em agricultura	58
16	Modelagem estrela do banco de dados	59

17	Processo de ETL modelado para inserir hierarquia regiões rurais na dimensão espacial	59
18	Processo de ETL modelado para inserir hierarquia fazenda na dimensão espacial	60
19	Processo de ETL modelado para inserir hierarquia talhões na dimensão espacial	61
20	Processo de ETL modelado para inserir fatos de rendimento, buscando as chaves de suas respectivas dimensões	61
21	Interface da implementação do método proposto: (a) Seleção de cubos e medidas disponíveis; (b) lista de dimensões dispostas em intensidade de cor; (c) mapa da dimensão espacial; (d) gráfico de barra da dimensão temporal; (e) gráfico de pizza da dimensão espacial; (f) Gráfico de barra da dimensão espacial; e (g) Tabela de medidas	63
22	Operações de <i>Drill-down</i> sucessivas e o comportamento dos gráficos e tabelas .	64
23	Operações de filtro para selecionar a produção de milho no ano de 2015, para os talhões da fazenda Paiquerê	65

Lista de Quadros

2.1	Vantagens relativas da representação <i>raster</i> e vetor	31
4.1	Análise das propostas de integração DW, SIG e AP	50
5.1	Arquivos da base de dados operacionais e seus respectivos campos	54

Lista de siglas

AGRIBI	<i>Agricultural Business Intelligence</i>
AP	Agricultura de Precisão
API	<i>Application Programming Interface</i>
BI	<i>Business Intelligence</i>
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CSS	<i>Cascading Style Sheets</i>
DBMS	<i>Database Management System</i>
DW	<i>Data Warehouse</i>
ETL	<i>Extract, Transformation, and Load</i>
FK	<i>Foreign Key</i>
GIS	<i>Geographic Information System</i>
GDW	<i>Geographical Data Warehouse</i>
ha	hectare
IBGE	<i>Instituto Brasileiro de Geografia e Estatística</i>
SGBD	Sistemas Gerenciadores de Banco de Dados
MDX	<i>MultiDimensional Expressions</i>
SVG	<i>Scalable Vector Graphics</i>
CSS	Cascading Style Sheets
OGC	<i>Open Geospatial Consortium</i>
OLAP	<i>Online Analytical Processing</i>
PRML	<i>Personalization Rules Modeling Language</i>
SDW	<i>Spatial Data Warehouse</i>
SGBD	Sistema Gerenciador de Banco de Dados
SIG	Sistemas de Informação Geográfica
SOLAP	<i>Spatial Online Analytical Processing</i>

SQL	<i>Structured Query Language</i>
UML	<i>Unified Modeling Language</i>

Sumário

1	Introdução	17
1.1	Objetivos	19
1.2	Estrutura da Dissertação	19
2	Sistemas de Suporte à Decisão em Agricultura	21
2.1	Introdução	21
2.2	Sistemas Agrícolas para Agricultura de Precisão	21
2.3	Data Warehouse	23
2.4	<i>Online Analytical Processing</i> (OLAP)	26
2.5	Sistemas de Informação Geográfica	28
2.6	<i>Spatial</i> OLAP (SOLAP)	31
2.7	Considerações	34
3	Tecnologias Abertas e Extensíveis Aplicadas ao Suporte a Decisão na Agricultura	35
3.1	Introdução	35
3.2	Etapas ETL	36
3.3	PostgreSQL e PostGIS	37
3.4	Geomondrian	38
3.5	OLAP4J e Mandoline	38

3.6	Geonode e Analytics	39
3.7	API de consulta	39
3.8	Crossfilter	40
3.9	Crossfilter server	41
3.10	DC.js e D3.js	42
3.11	Considerações	42
4	Requisitos de Integração e Trabalhos Correlatos	43
4.1	Introdução	43
4.2	Requisitos de Integração	43
4.3	Trabalhos Correlatos à <i>Data Warehouse</i> espacial e agricultura	45
4.3.1	Personalização para um <i>Data Warehouse</i> espacial	46
4.3.2	Spatial Data Warehouse e Spatial OLAP	46
4.3.3	Geographical Data Warehouses	47
4.3.4	EIS Pesticides	48
4.4	Abordagem Comparativa	49
4.5	Considerações	49
5	Material e Métodos	51
5.1	Introdução	51
5.2	Métodos	51
5.2.1	Fonte de dados operacionais	53
5.2.2	Processos ETL	53
5.2.3	Banco de dados Multidimensional	54
5.2.4	API SOLAP	54
5.2.5	Aplicação	55
5.2.6	Interface	55
5.3	Considerações	55

6	Resultados e Discussões	56
6.1	Arquitetura AGRIBI	56
6.1.1	Banco de dados	57
6.1.2	Aplicação	60
6.1.3	Interface	62
6.2	Considerações	64
7	Considerações Finais	66
	REFERÊNCIAS	68
	APÊNDICE A – Instalação de bibliotecas e módulos	72
A.1	Exportar arquivos do software APEX	73
A.2	Instalação do SGBD PostgreSQL e a extensão PostGIS	74
A.3	Instalação e configuração GeoMondrian	75
A.3.1	Pré requisitos	75
A.3.2	Instalação	75
A.3.3	Configuração	75
A.4	Instalação API SOLAP	77
A.5	Instalação Geonode	78
A.6	Instalação da extensão analytics e configuração	78

Introdução

Agricultura de precisão se caracteriza pelo gerenciamento da área plantada, mediante sistemas de informação e equipamentos que permitem a aplicação do insumo, no momento adequado e na área afetada (FRAZIER; WALTERS; PERRY, 1997). A variabilidade presente no solo resulta em diferentes resultados de produtividade, assim como diferentes manejos para cada cultura.

O advento das tecnologias para obtenção de dados espacialmente referenciados e Sistemas de Informação Geográfica tornaram possível a coleta e o gerenciamento de informações relativas a um ponto específico da área cultivada.

De acordo com McBRATNEY, WHELAN e ANCEV (2005), a produtividade é influenciada por propriedades do solo como umidade, teor de argila, teor de matéria orgânica, nutrientes, pH e densidade do solo. As variações podem ser de natureza sazonais, tais como estresse hídrico, falta de nutrientes, doenças, ervas daninhas, pragas, entre outras. Os valores dessas propriedades são coletados por meio de amostras, por unidade de área, e suficientes para representar a variabilidade na área agrícola.

A obtenção de propriedades, por meio de amostras, pode ser dispendiosa. Uma alternativa é o sensoriamento remoto (FRAZIER; WALTERS; PERRY, 1997). A detecção remota pode avaliar as condições de cultivo por meio da refletância da superfície, do solo, das plantas ou da mistura de ambos.

O sensoriamento remoto não supre todas as demandas de informações sobre o cultivo, tais como dos dados referentes ao sistema radicular da planta. Estas tecnologias podem ser utilizadas em conjunto, para obtenção de um maior número de informações.

As informações obtidas sobre a área cultivada são indicadores de produtividade utilizados para avaliar o desempenho da cultura. Os mapas de produtividade resumem o impacto das condições naturais, do clima e do solo, e permitem o gerenciamento de manejos na aplicação de adubação, na densidade de plantas, na aplicação de defensivos agrícolas e na rotação de culturas, para o melhor aproveitamento do potencial produtivo específico de cada ponto da área produtiva.

Uma abordagem preditiva pode ser adotada para o gerenciamento do ciclo produtivo. As informações são armazenadas ao longo do tempo, de forma a produzir um histórico por meio de mapas temáticos, topografia, e por outras entradas espaciais. Essas informações podem ser utilizadas para prever o tratamento e manejo adequado para eliminar fatores limitantes de produtividade, tais como baixo pH ou compactação do solo, com taxas variáveis de aplicação de insumos. Isso ocorre como medidas corretivas para obter um melhor desempenho da cultura. Diferentemente, devem ser tratados fatores imutáveis ou quando o custo/benefício é inviável, com baixa capacidade de retenção de água e impossibilidade de irrigação, deve ser reduzida a quantidade de insumos aplicada, uma vez que não consumida pela cultura e desperdiçada no ambiente (HEEGE; REUSCH; THIESSEN, 2008).

Frente à necessidade de gerenciamento e tomadas de decisão, durante o ciclo da agricultura de precisão, faz-se necessário o uso de sistemas de suporte a decisão. O objetivo dos sistemas de suporte à decisão é gerenciar e analisar, de forma eficiente e consistente bases de dados, para extrair informações, identificar tendências, padrões e gerar prognósticos. Dentre as ferramentas de suporte a tomada de decisão, esta pesquisa foca o uso de *Data Warehouse - DW*, ferramenta para processamento de dados multidimensional (*On-Line Analytical Processing - OLAP*) e processamento de dados geográficos (Sistema de Informação Geográfica - SIG).

Um Data Warehouse é considerado uma base de dados modelada e populada para fornecer dados inerentes à tomada de decisão. Já uma Ferramenta OLAP é responsável pelo processamento de consultas multidimensionais sobre um Data Warehouse. Um Sistema de Informação Geográfica é especializado em adquirir, gerenciar, manipular e analisar dados espacialmente referenciados.

No âmbito da Agricultura de Precisão, há demanda de contexto multidimensional e de análise dos dados sobre o contexto espacial. Assim, a integração das Ferramentas DW, OLAP e SIG tem um papel relevante na tarefa de responder questões como “o que?”, “quem?”,

“quando?”, “onde?”. Por exemplo, dado uma determinada área agrícola, (i) onde é a área que o custo de produção de soja é maior que R\$1,800.00 por ha (hectare)?; (ii) qual o local que teve a maior produtividade do milho entre as safras de 2010 e 2014?; (iii) qual a quantidade de inseticida aplicado nos últimos 5 anos dentro de uma área definida?

A integração entre DW e OLAP, com as funcionalidades do SIG, para dados agrícolas na Agricultura de Precisão, não é trivial. O propósito de cada um é diferente, e seus dados, metadados e serviços não são diretamente interligados, e as informações agrícolas necessitam de um tratamento especial, na questão de sensores com calibração inadequada.

A contribuição desta pesquisa é uma arquitetura aberta, integrada, com uma modelagem de *data warehouse* agrícola espacial, auxiliando a tomada de decisão para o planejamento e gestão do manejo das práticas da agricultura de precisão.

1.1 Objetivos

O objetivo desta pesquisa é definir uma arquitetura para apoio à decisão em Agricultura de Precisão, utilizando as ferramentas integradas de *Data Warehouse* e Sistemas de Informação Geográfica.

Os objetivos específicos são os que seguem: (i) analisar frameworks de implementação de DW/OLAP, SIG e AP, para proporcionar a integração; (ii) modelar um sistema agrícola para dar suporte à tomada de decisão, com base em dados espacialmente referenciados; (iii) analisar padrões e protocolos de comunicação; (iv) avaliar a arquitetura para implementação da integração DW/OLAP, SIG e AP; (v) avaliar camadas de apresentação interface gráfica para visualização de consultas e respostas; (vi) Utilizar padrões e arquiteturas abertas, para redução de custos, fácil aplicação e extensibilidade; e (vii) reusar soluções parciais de modelagem multidimensional e geográfica, sistemas de apoio a decisão em agricultura de precisão e dados espacialmente referenciados.

1.2 Estrutura da Dissertação

Afim de apresentar os conceitos e tecnologias abordadas, bem como os objetivos alcançados, o presente documento está organizado da seguinte forma:

Capítulo 2 - Sistemas de Suporte à Decisão em Agricultura: Apresenta conceitos relacionados a sistemas agrícolas e de suporte a decisão. Os tópicos abordados se referem a *Sistemas para agricultura de precisão, Data Warehouse, OLAP, Sistemas de Informação Geo-*

gráfica e *Data Warehouse* Geográfico.

Capítulo 3 - Tecnologias Abertas e Extensíveis Aplicadas ao Suporte a Decisão na Agricultura: Este capítulo é dedicado a apresentar as tecnologias abertas e extensíveis que possam ser utilizadas para implementação dos conceitos de sistemas de suporte a decisão.

Capítulo 4 - Requisitos de Integração e Trabalhos Correlatos: São citadas literaturas que estão correlacionadas com a presente dissertação, considerando pesquisas relacionadas a modelagem multidimensional com dimensões espaciais, comparando-as com os seguintes critérios: camada de apresentação, modelagem de dados agrícolas, modelagem espacial, modelagem multidimensional, integração com AP e linguagem de consulta.

Capítulo 5 - Material e Métodos: Este capítulo descreve os procedimentos realizados para o projeto e implementação da integração proposta.

Capítulo 6 - Resultados e Discussões: Este capítulo é dedicado a apresentar os resultados obtidos a partir da integração proposta, sendo a arquitetura, processo ETL e modelagem agrícola e dimensional.

Capítulo 7 - Considerações Finais: E por fim este capítulo destaca as conclusões obtidas por meio do desenvolvimento desta pesquisa, e os trabalhos futuros relacionados.

Sistemas de Suporte à Decisão em Agricultura

2.1 Introdução

Segundo Fountas *et al.* (2015), melhorias são necessárias no quesito de sistemas de suporte à decisão, para realizar a gestão do conhecimento da agricultura, abordando aspectos geoespaciais e temporais ao longo do ciclo produtivo. Para o atendimento à demanda de sistemas de armazenamento, gerenciamento e de recuperação de dados multidimensionais e espaciais, é necessária uma arquitetura de software e modelagem de dados agrícolas adequadas para atendimento o contexto da Agricultura de Precisão (NIKKILA; SEILONEN; KOSKINEN, 2010).

2.2 Sistemas Agrícolas para Agricultura de Precisão

Para o desenvolvimento de sistemas de apoio à decisão agrícola, os requisitos não-funcionais são os que seguem (MURAKAMI *et al.*, 2007): (i) atender as necessidades específicas do agricultor; (ii) disponibilizar funcionalidades simples e automatizadas para o processamento de dados; (iii) ter o controle dos parâmetros de processamento e funções de análise, assim como permitir novas soluções; (iv) flexibilizar sistemas especialistas, com conhecimento baseado em regras, para inclusão da experiência do agricultor, práticas e preferências; (v) possuir padrões de integração para redução de custos, curva de aprendizagem e suporte; (vi) permitir a interoperabilidade entre os módulos e fontes de dados; (vii) permitir a escalabilidade para diferentes

necessidades; e (viii) ser de baixo custo. Com esses requisitos, é necessária a modelagem dos sistemas, por meio da identificação e definição do escopo e fronteiras.

Uma vez observados os requisitos do sistema é necessária a modelagem conceitual com a definição dos atores e funcionalidades. O modelo conceitual pode ser dividido em quatro seções, envolvendo conjunto de dados internos, coleção de informações externas, planejamento e geração de relatórios (MURAKAMI *et al.*, 2007).

A coleta de dados e o processamento são realizados de forma automatizada, enquanto os relatórios e planejamento devem ser inicializados e geridos pelo gestor da área agrícola. Já as informações externas provêm de padrões, regras, e normas de funcionamento das atividades agrícolas (SORENSEN *et al.*, 2011).

No fluxo da informação agrícola, os dados podem ser classificados nos níveis estratégico, tático e operacional. Os atores e os processos de decisão são definidos, em cada operação agrícola. As operações podem ser definidas como plantio, semeadura, adubação, pulverização, irrigação e colheita e são parâmetros para a análise e projeção, para tomada de decisão em nível tanto tático como estratégico (SORENSEN *et al.*, 2011).

A tomada de decisão abrange processos (Figura 1), como visão futura da produção, planejamento geral do cultivo, elaboração da estratégia de aplicação de insumo, seleção das tecnologias de informação e máquinas, seleção das características do insumo, obtenção das informações para planejamento, planejamento, inspeção da área agrícola, atualização do plano de execução, seleção de parâmetros, inspeção e controle da execução do manejo planejado, finalização ou reformulação do plano executado, documentação, resumo do desempenho e comparação do planejado com o realizado (GEBBERS; ADAMCHUK, 2010).

Processos de decisão estão fortemente interligados à necessidade de obter informações. Dados da cultura, como vigor e presença de doenças, solo, umidade e acidez, podem ser obtidos por meio de amostragem e análise laboratorial, ou por levantamento visual. Sistemas de sensoriamento avaliam diretamente a variância espacial através da refletância espectral, para obtenção das condições da cultura, como índice de área foliar, cobertura do solo e índice de nitrogênio.

A complexa interação entre o solo e a cultura dificulta a definição do tratamento variável dos fatores do solo. Entretanto, o mapeamento das condições do solo permite identificar fatores limitantes, como o mapeamento de deficiências nutricionais, podendo indicar as áreas de deficiência, mas não ser diretamente utilizado para a aplicação de fertilizantes (STAFFORD, 2000).

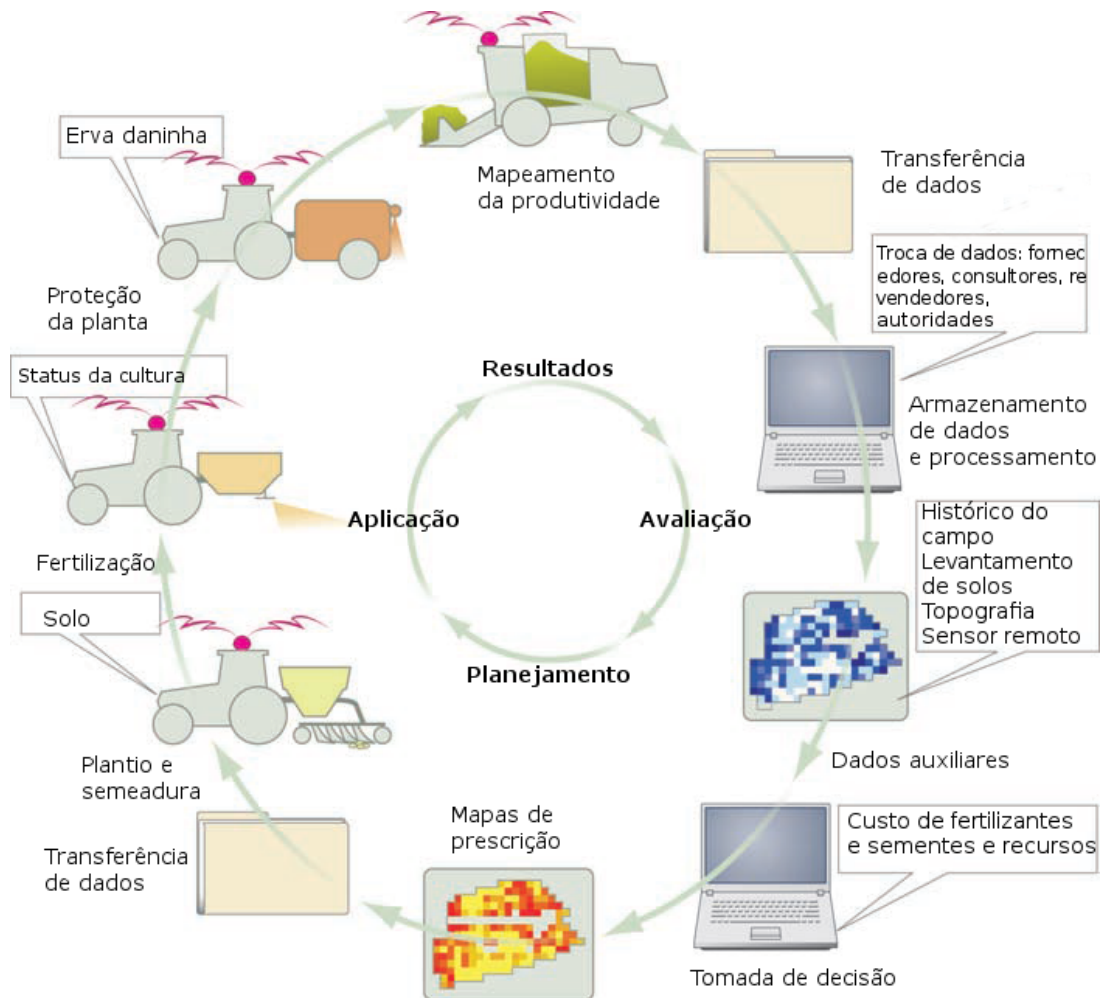


Figura 1: Fluxo da informação do ciclo produtivo agrícola, aplicando processos de análise e manejo na agricultura de precisão.

Fonte: (GEBBERS; ADAMCHUK, 2010)

Os dados agrícolas obtidos em campo, seja por meio de sensores remotos, ou por análises, são armazenados em bases de dados. Estas por sua vez são classificadas como operacionais e orientados a transação, seguindo a modelagem relacional (ELMASRI; NAVATHE, 2010). Já uma base de dados para suporte à decisão possui um modelo multidimensional, como o modelo estrela, e é orientado por assunto (KIMBALL; ROSS, 2013)). Dessa maneira, a arquitetura de dados deve ser dividida em dois ambientes, um ambiente operacional e outro para suporte à decisão.

2.3 Data Warehouse

Para a construção do ambiente de suporte à decisão, utiliza-se o conceito de *Data Warehouse* (DW). As informações de um DW são (KIMBALL; ROSS, 2013): (i) facilmente acessíveis, devendo ser óbvias para o usuário; (ii) consistentes, originadas de várias fontes e devem ser montadas; (iii) adaptáveis a mudanças, visto que a regra de negócio sofre mudanças e o

sistema deve ser flexível; (iv) variantes no tempo, permitindo histórico de informações; (v) seguras, sendo que os dados devem ser tratados de maneira confidencial e com controle de acesso; e (vi) aceitáveis pelo usuário gerenciador.

A arquitetura de um ambiente DW pode ser dividida em quatro componentes (Figura 2): fonte operacional, Sistema ETL (*Extract, Transformation, and Load*), dados na área de apresentação e aplicações de suporte à decisão. A camada de aplicações é composta por consultas ad hoc, padrões de relatórios, aplicativos analíticos e mineração de dados, porém, neste trabalho serão abordados apenas aplicativos analíticos.

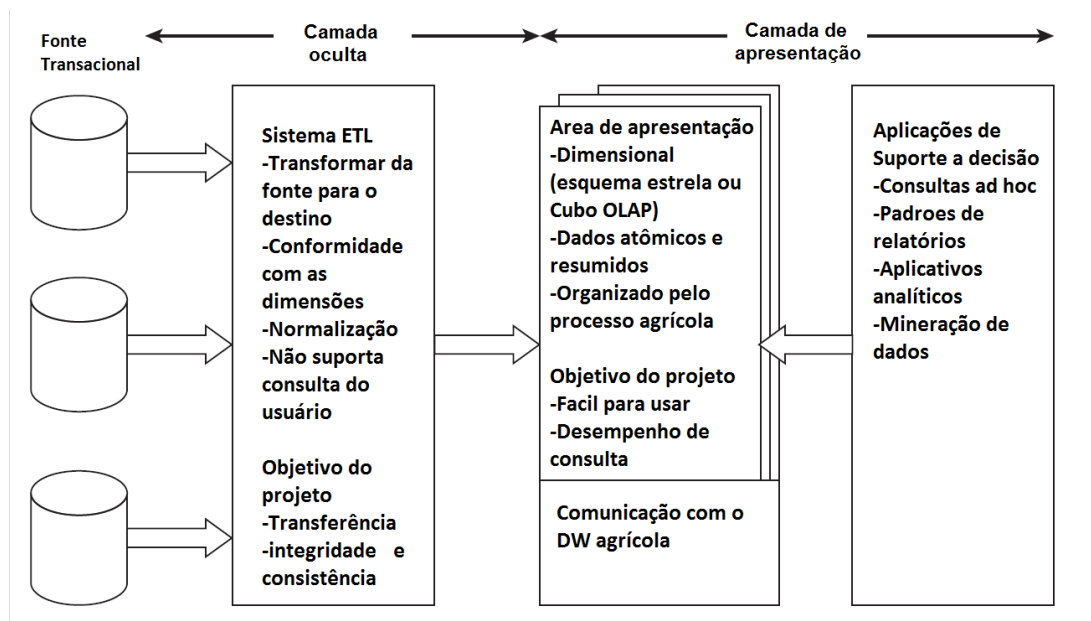


Figura 2: Arquitetura de um Data Warehouse e suas camadas
Fonte: Baseada em Kimball e Ross (2013)

O Sistema ETL de um DW consiste em área de trabalho, estruturas de dados instanciadas e no set de processos, meio de comunicação entre a fonte operacional e a área de apresentação dos dados no DW. O primeiro passo consiste na leitura e compreensão dos dados da origem operacional para o DW. O segundo passo é a limpeza dos dados, como correção de erros de ortografia, conflito de domínios, falta de elementos e combinando dados de várias fontes. A última etapa é a estruturação física e carregamento dos dados no modelo dimensional da camada de apresentação (KIMBALL; ROSS, 2013).

Na área de apresentação de um ambiente DW, pode ter o componente *Data Mart*. O *Data Mart* é semelhante ao DW, mas possui escopo restrito a um assunto da área agrícola, como setor de fertilizantes. E por ser restrito, o *Data Mart* tem um menor tempo de implementação e uma menor utilização de recursos (KIMBALL; ROSS, 2013).

Um DW (ou um *Data Mart*) é modelado a partir de modelos dimensionais, que podem

ser modelo estrela ou cubo (Figura 3). O Modelo Estrela é implementado em um SGBD relacional e é conhecido pela sua estrutura se assemelhar a uma estrela. Já o cubo é implementado em um banco de dados multidimensional. Ambos modelos possuem uma lógica de modelagem semelhante, baseado nas dimensões identificadas, mas a implementação física difere. O cubo possui vantagem em desempenho, por utilizar pré-processamento, agregações, estratégia de indexação e funções analíticas robustas, disponíveis na Linguagem SQL (Structured Query Language) (KIMBALL; ROSS, 2013).

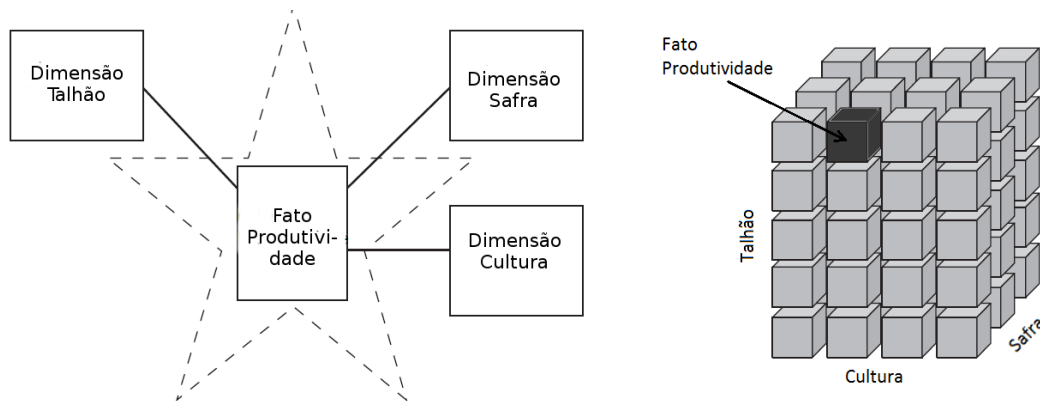


Figura 3: Representação gráfica do Modelo estrela e cubo de um Data Warehouse

A tabela de fatos, no modelo dimensional, é responsável pela medida de desempenho da área agrícola. Todos os dados do processo agrícola devem resultar em um único modelo dimensional, para evitar replicações e inconsistências.

O termo fato é uma medida de desempenho do processo agrícola, como colheita de uma cultura de grãos, com cada passo que a máquina colhedora realiza, registrado no banco de dados operacional, envolvendo dados de velocidade, umidade e quantidade. Por meio dos dados operacionais, é possível gerar a tabela de fatos, descrevendo os atributos para a medida de desempenho da área agrícola (Figura 4).

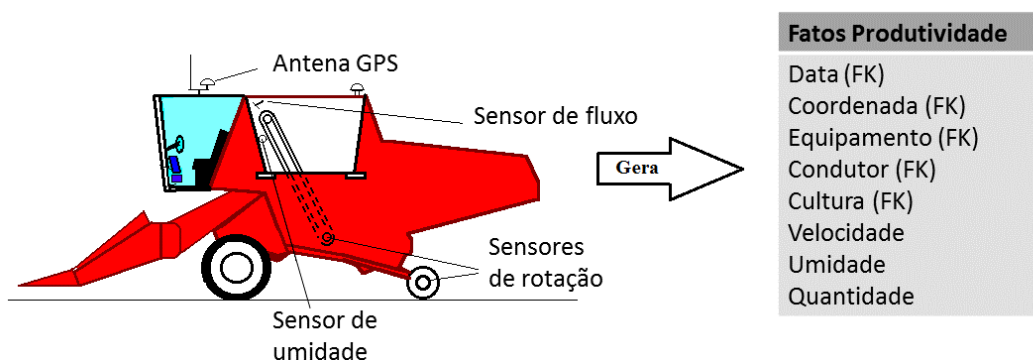


Figura 4: Esquema de obtenção das medidas de produtividade do processo produtivo através dos sensores da colhedora

A tabela de fatos (Figura 4) é composta pelas chaves estrangeiras, ou Foreign Key

(FK) que identificam as dimensões, as quais os fatos estão ligados, e os próprios fatos que são coletados em cada evento de mensuração.

A tabela de fatos contém fatos numéricos e aditivos, mas também podem ser encontrados fatos semi-aditivos, e não aditivos. As entradas aditivas correspondem a valores contínuos, adicionados em todas as dimensões e quantificando a atividade, como a quantidade de grãos colhida. Os fatos semi-aditivos não são registrados em algumas dimensões, e correspondem às medidas de intensidade, como velocidade da máquina. Já fatos não aditivos representam valores textuais, que podem ser contados ou apenas mostrados um-a-um, como descrição da qualidade dos grãos (bom, médio ou ruim). Porém, este último tipo de fato não é recomendado, pois informações textuais apenas permitem contagem e estatísticas associadas a contagens (KIMBALL; ROSS, 2013).

As tabelas de dimensão fornecem a descrição do contexto do processo agrícola, associado a um evento de mensuração. Deve descrever “quem, o que, onde, quando, como e porque”. Normalmente as tabelas de dimensão tendem a conter muitos atributos, e poucas linhas, se comparadas a tabela de fatos. Cada dimensão é composta por uma chave primária que irá identificar unicamente a dimensão, e os demais atributos são denominados membros, organizados e agregados em níveis hierárquicos.

2.4 *Online Analytical Processing (OLAP)*

Online Analytical Processing (OLAP) é um conceito de processamento multidimensional dos dados, realizado em um modelo como por exemplo o modelo estrela. Independente destes modelos, as ferramentas OLAP manipulam e extraem dados do DW em estruturas lógicas denominadas cubos de dados.

Os operadores OLAP permitem a manipulação da estrutura do cubo, permitindo a navegação ao longo das hierarquias conceituais, combinações entre dimensões e visualização de resultados. As consultas por meio de operadores podem ser realizadas em bases de dados multidimensionais pela linguagem MDX (*Multidimensional Expressions*). Por meio da sintaxe MDX pode-se realizar consultas em um cubo de dados multidimensional para obter visões configuráveis dos dados em diferentes ângulos e níveis de agregação. Essas consultas podem ser realizadas através da aplicação de operadores como *slice/dice*, *rotate* ou *pivot* e *drill-down/roll-up*. O operador *slice/dice* permite fatiar o cubo multidimensional de acordo com um filtro em uma das dimensões. A operação de *rotate* ou *pivot* realiza uma rotação nos eixos do cubo de dados, provendo uma visão alternativa dos dados. E as operações de *drill-down/roll-up* diminuem ou aumentam o nível de granularidade da informação (FU; HAMMER, 2001).

Em alguns aspectos a sintaxe MDX é muito semelhante a linguagem SQL, como por exemplo as construções *select*, *from* e *where* e algumas funcionalidades suportadas pela MDX podem ser fornecidas pela SQL (MELTON; SIMON, 1993). Porém a sintaxe MDX não é uma extensão da SQL e esta não é eficiente e intuitiva para os cenários multidimensionais (RAO, 1996).

Para a navegação das hierarquias conceituais, utiliza-se o *roll-up* para a agregação (ou generalização) dos dados, permitindo a análise em cenários mais sumarizados e o *drill-down* para a desagregação (ou especialização) dos dados, tornando possível avaliar os dados de forma mais detalhada (THOMSEN, 2002). Por exemplo, ao ser analisado os dados de produtividade no Brasil, uma operação de *drill-down* na dimensão espacial na região dos campos gerais, resultaria nos dados das cidades que compõem a região dos campos gerais (Ponta Grossa, Castro, Palmeira, etc). Um novo *drill-down* pode ser realizado em cada uma das cidades retornadas pela operação anterior, sendo que esta operação retornará o próximo nível da hierarquia da dimensão espaço, que pode ser por exemplo as fazendas cadastradas correspondente a cidade a qual foi selecionada.

Para modificar a orientação do cubo é utilizada a operação *rotate* ou *pivot* que troca a posição das dimensões entre os eixos do cubo gerando uma nova configuração de análise dimensional. Para exemplificar melhor a operação *rotate*, se formatarmos o resultado de uma consulta OLAP em uma tabela, esta operação "rotaciona" a tabela trocando linhas por colunas e vice-versa.

Já o processo de combinar dimensões, tem-se a operação *slice* para a extração de um sub-cubo das células, verificando restrições em uma dimensão, e a operação de *dice*, que permite a extração de um sub-cubo com restrições em duas ou mais dimensões. Nestas operações o foco da análise é para uma camada de dados particular, fatiando um subconjunto de dados agregados (THOMSEN, 2002). Essas operações podem ser ilustradas na Figura 5, modificada a partir da figura de Paim (2003) para um exemplo no contexto deste trabalho.

A Figura 5 apresenta uma representação gráfica para as operações OLAP, de forma a interpretar os dados como um cubo, onde cada uma das dimensões do *data warehouse* são representadas pelas dimensões físicas do cubo, ilustrando a operação de *slice* (fatiando o cubo), *dice* (retirando um subcubo), *Drill-down* (subdividindo um subcubo), *Roll-up* (agregando em um cubo generalizado) e a operação de *rotate* (troca de dimensões).

Neste capítulo foram abordadas as operações que podem ser realizadas sobre as informações de forma analítica. No próximo capítulo será abordados, sistemas de informações geográficas, e a forma de representação de informações geométricas.

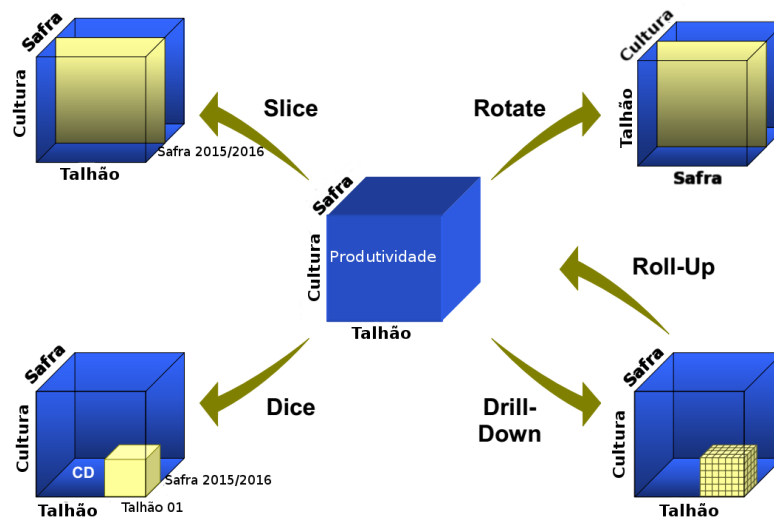


Figura 5: Esquema de exemplos e suas representações gráficas das operações OLAP em um cubo

2.5 Sistemas de Informação Geográfica

Para o suporte, as características espaciais utiliza-se os Sistemas de Informação Geográfica (SIG) (DEMERS, 2008). Um dado espacial é composto por atributos gráficos e descritivos, possuindo quatro componentes para descrever um objeto do mundo real: a posição geográfica, os atributos, as relações espaciais e o tempo (BURROUGH, 1986).

A inspeção das feições geográficas é realizada visualmente por meio de mapas, que é uma representação abstrata da superfície terrestre. Para a representação das diversas informação geográfica, como umidade do solo, relevo, tipos de solo, matéria orgânica e nutrientes, as ferramentas SIG possuem o recurso de sobreposição de camadas. Assim, podem ter diversos temas de informações georreferenciadas combinados para o processamento de uma consulta (Figura 6).

Em um SIG a coleta e armazenamento de dados é uma fase crítica, complexa e onerosa (BURROUGH; MCDONNELL, 1998). Segundo Paulino e Carneiro (1998), a origem de dados espaciais pode ser dividida em três categorias: (i) dados provenientes de outros sistemas; (ii) dados provenientes de bases analógicas; e, (iii) dados adquiridos de fontes primárias.

Um dado provenientes de sistemas externos é relacionado à processos de migração e conversão de bases de dados e ocorre quando há incompatibilidades entre softwares ou estruturas de dados (nova modelagem).

Já as bases analógicas é comum na aquisição manual ou baseada em processos de digitação, "escanerização" e vetorização (automática, semi-automática ou manual).

A última categoria refere-se a métodos de produção de dados espaciais, em formato

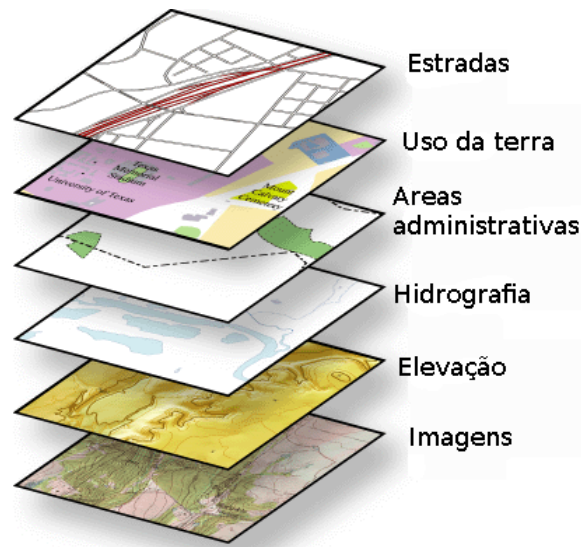


Figura 6: Sobreposição de camadas para representar níveis de informação sobre uma determinada área

digital, a partir de dados primários, ou seja, sensoriamento remoto por meio de satélites, imageamento por meio de aeronave remotamente pilotada (métodos indiretos) ou levantamentos geodésicos (métodos diretos).

Como definido, um dado espacial é de natureza gráfica e descritiva. Portanto precisam de estruturas específicas para armazenar cada uma das duas naturezas. Assim como procedimentos e funções específicas para a recuperação, processamento e exibição.

A natureza gráfica poder ser representada por meio de uma estrutura matricial ou vetorial. Na estrutura matricial, também conhecida como *raster*, discretiza-se a superfície em forma de matrizes, sendo o elemento básico e indivisível a célula ou pixel (do inglês, *picture element*) (ARAÚJO, 2003). Esta representação supõe uma superfície plana para tratar o espaço onde cada célula é associada a uma porção do terreno, podendo definir linhas, áreas ou pontos (CÂMARA *et al.*, 1996). A Figura 7 mostra um exemplo de representação de um espaço utilizando a representação matricial da natureza gráfica do dado.

O modelo vetorial de representação, trata a superfície de forma contínua através de coordenadas e seguindo três primitivas geométricas básicas: ponto, linha e polígono. A Figura 8 apresenta exemplos da representação vetorial, onde pode-se observar que o ponto é definido por um único par de coordenadas. A linha é definida por um conjunto de pontos, sendo que o ponto inicial e final são os nós. E o polígono é um conjunto de linhas que se conectam definindo uma área.

As duas formas de representação possuem suas vantagens e desvantagens. O quadro 2.1 apresenta um resumo da visão de ARONOFF (1989) sobre a questão.

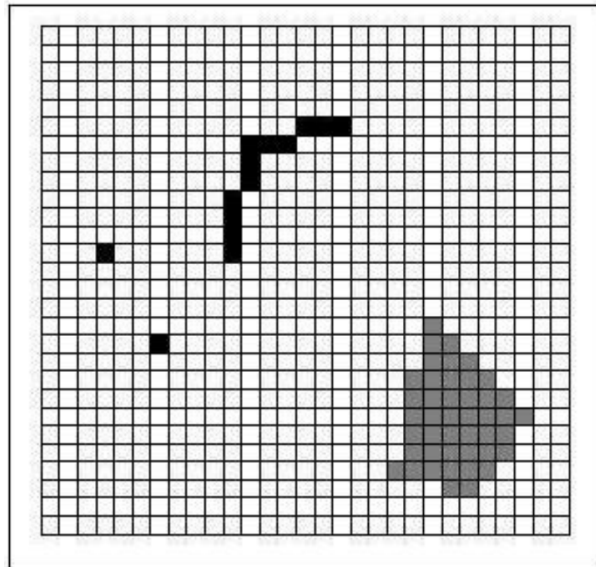


Figura 7: Esquema de representação espacial da estrutura matricial

Fonte: (ARAÚJO, 2003)

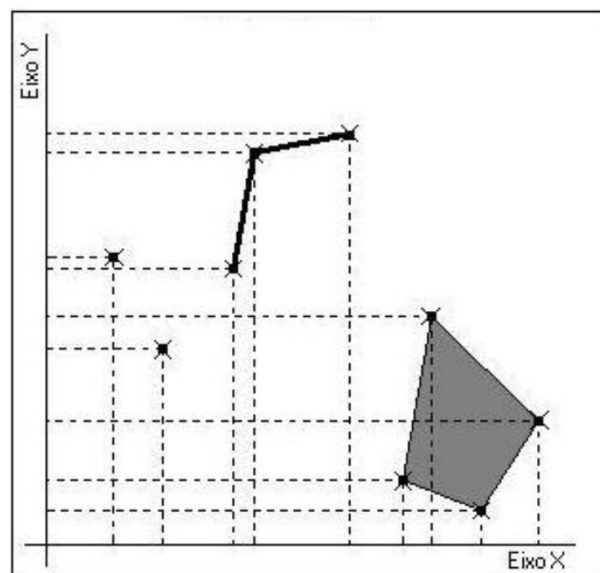


Figura 8: Esquema de Representação Espacial da Estrutura Vetorial

Fonte: (ARAÚJO, 2003)

Os dados, tanto vetorial como matricial, podem ser manipulados de várias formas. Os procedimentos fazem parte das especificações de recursos do OGC, dos quais os principais métodos de relações espaciais são: *Equals*, *Disjoint*, *Intersects*, *Touches*, *Crosses*, *Within*, *Contains*, *Overlaps* e *Relate*. Já os métodos de análise espacial estão entre os principais: *Distance*, *Buffer*, *Convex Hull*, *Intersection*, *Union*, *Difference* e *Symmetric difference* (SMITH; GOODCHILD; LONGLEY, 2015).

As características espaciais, como dimensão, posição, tamanho, forma e orientação, em um sistema computacional devem ser abstraídas por meio de um objeto geométrico. O mundo

Quadro 2.1: Vantagens relativas da representação *raster* e vetor

	Matricial	Vetorial
Vantagens	1. estrutura de dados mais simples; 2. operações entre níveis de informação são mais facilmente e eficientemente implementadas; 3. fenômenos com alta variabilidade espacial são representados mais eficientemente; e 4. apropriado para utilização de imagens digitais.	1. fornece uma estrutura de dados mais compacta (menos espaço para armazenamento); 2. trabalha eficientemente com topologia, e por consequência, fornece melhores resultados para operações que requerem informação topológica, como análise de redes; e 3. é mais apropriado a suportar gráficos que se aproximam de mapas desenhados à mão.
Desvantagens	1. maior espaço de armazenamento; 2. as relações topológicas são mais difíceis de representar; e 3. a saída é menos estética.	1. a estrutura de dados é mais complexa; 2. as operações entre níveis de informação são mais difíceis de implementar; e 3. a representação de alta variabilidade espacial é ineficiente.

Fonte: Adaptado de (ARONOFF, 1989)

é representado por um conjunto de objetos identificáveis e localizáveis, onde cada objeto tem suas características próprias e uma geometria associada, definindo seus limites no espaço (OGC, 2007). De acordo com OGC, cada objeto geográfico pode ser definido por meio das especializações da classe *GM_Object* (Figura 9).

Na Figura 9 a classe *GM_object* é a generalização de todos os tipos de objetos geométricos. A classe especializada *GM_Primitive* é a classe genérica para os objetos únicos, orientados e homogêneos, como pontos, curvas, superfície e sólidos. Já a classe *GM_Complex* corresponde a uma composição de geometrias primitivas (*GM_Composite*). Finalmente, a Classe *GM_Aggregate* representa uma coleção de objetos sem estrutura interna, objetos geográficos sem relacionamentos definidos. As subclasses devem ser relacionadas aos tipos primitivos (*GM_MultiPrimitive*), mas diferentemente dos tipos primitivos, estas podem ser decompostas em multipontos, multicurvas, multisuperfícies e multisólidos (OGC, 2007).

2.6 Spatial OLAP (SOLAP)

Spatial OLAP (SOLAP), ou OLAP espacial é um termo utilizado para definir um sistema que utiliza as tecnologias de consulta multidimensional e multinível em conjunto com

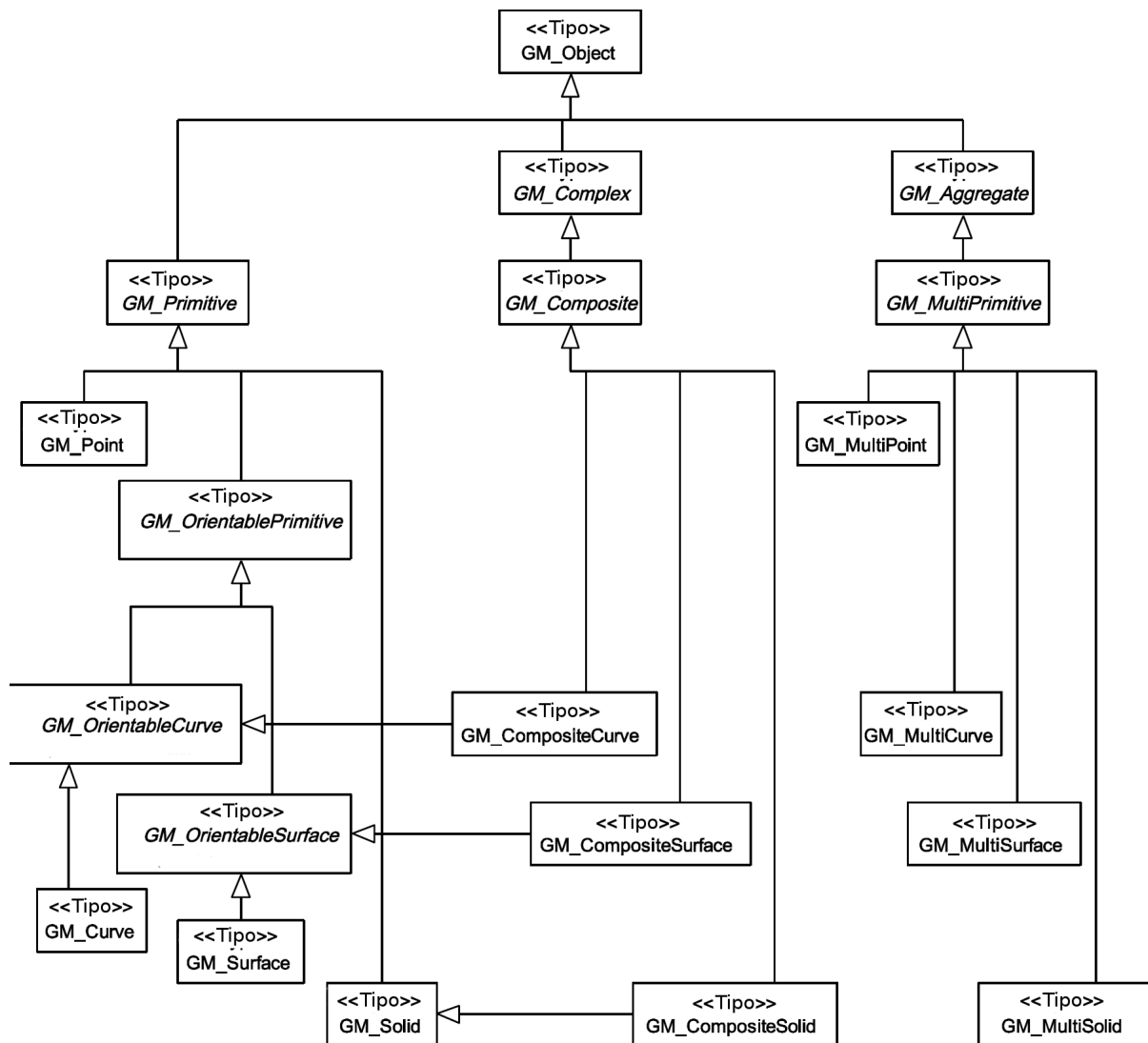


Figura 9: Diagrama de classe - especialização GM_Object (OGC, 2007)

tecnologias SIG (BÉDARD; MERRETT; HAN, 2001). Portanto, o intuito da junção destas duas tecnologias de suporte a decisão é proporcionar ao utilizador a flexibilidade de consultas OLAP em um *data warehouse* e a correta manipulação e recursos gráficos dos atributos geográficos, ou seja, da geolocalização da informações (INMON, 1995).

A arquitetura de software para prover um sistema SOLAP deve ser composta pelas seguintes camadas (GOMES, 2011) (BIMONTE; WEHRLE, 2006): (i) Camada de armazenamento: Composta por um DWG (*Data Warehouse Geográfico*) para armazenar dimensões (com atributos geográficos ou não) e os fatos a serem sumarizados através das dimensões geográficas e não geográficas; (ii) Camada do servidor: contém os serviços para prover as operações OLAP e SIG, de forma integrada afim de analisar dados geográficos e multidimensionais. Pode ser composta por um servidor OLAP (exemplo Pentaho Mondrian (HYDE, 2016) e MS Analysis

Service (MICROSOFT, 2016)) integrado a um serviço espacial (exemplo Oracle espacial (ORACLE, 2016b) e PostGIS (OSGEO, 2016); e (iii) Camada cliente: Interface gráfica que permite a visualização, exploração e análise de mapas, tabelas, gráficos. É formada por um cliente OLAP (exemplo JPIVOT (TONBELLER, 2016)) integrado a uma ferramenta de visualização cartográfica (exemplo Oracle MapViewer (ORACLE, 2016a) ou Google Maps (GOOGLE, 2016)).

As operações de um OLAP espacial são consultaa multidimensionais com opções de filtrar, agregar, selecionar atributos geográficos (GOMES, 2011). Assim Gomes (2011) propõe as seguintes operações:

- **Projeção SOLAP:** Projeta no mapa um conjunto de níveis geográficos, por exemplo, "média de produtividade da soja no ano de 2015 por talhão, fazenda, região". A consulta apresenta informações descritivas (produtividade, operador, máquina, umidade) tabuladas ou gráficas e informações espaciais (geométricas) projetadas em camadas do mapa;
- **Seleção SOLAP:** Operações topológicas ou métricas sobre uma camada geográfica para a seleção de informações. Por exemplo "qual a produtividade do milho nos talhões disjuntos a cidade de Ponta Grossa e talhão distante até 10 km da sede da fazenda". Neste caso tem-se as operações "talhões disjuntos a cidade de Ponta Grossa" e "distante até 10 km da sede da fazenda";
- **Junção SOLAP:** realiza uma operação espacial sobre dois níveis geográficos. Por exemplo: Considerando os níveis "talhão" e "fazenda", qual a produtividade por safra e variedade, na colheita do trigo?

Especificando as operações de Projeção SOLAP e Seleção SOLAP, pode-se especializá-las em:

- **Projeção SOLAP de Navegação:** Pode-se realizar uma operação de para um nível geográfico acima (Roll-Up espacial) ou abaixo (Drill-Down espacial) de uma informação geográfica. Por exemplo: "Qual a máxima produtividade no ano de 2015 por fazenda". Fazendo um Drill-Down na Fazenda X, obtemos a mesma resposta por talhões da Fazenda X. Ou fazendo um Roll-Up temos a mesma resposta por cidades; e
- **Seleção SOLAP geométrica:** a consulta é realizada utilizando um parâmetro geométrico (Ponto, MultiPonto, Poligono, Multipoligono), podendo ser um ponto, uma região, ou até uma janela. Por exemplo "qual a produtividade média do talhão que contém o ponto P". Ou "qual a produtividade máxima da safra de 2015 de milho que está contido na janela J".

A característica geral de um sistema SOLAP são as operações multidimensionais com atributos geométricos, ou seja integração dos recursos presentes em um SIG com Data Warehouse multidimensional para suporte a decisão. Essas consultas multidimensionais e espaciais não podem ser processadas, individualmente por ferramentas SIG ou por ferramentas OLAP (BARROS, 2009).

2.7 Considerações

Neste capítulo foi apresentado os conceitos relativos ao suporte a decisão, como sistemas para agricultura de precisão, *Data Warehouse*, operações OLAP e ferramentas SIG. Por fim foi apresentado os conceitos que integram uma modelagem multidimensional de dados com atributos geográficos (SOLAP). O suporte a decisão em agricultura deve ser apoiado por sistemas que tangem aspectos analíticos e geográficos.

O processamento de um ambiente de suporte a decisão deve suportar a modelagem multidimensional e espacial, devido a natureza dos dados. Portanto a arquitetura deve suportar as operações OLAP tradicionais de um *Data Warehouse* como, *Drill-Down*, *Roll-up*, *Slice*, *Dice*, assim como suportar atributos geográficos e dimensões baseadas nestes atributos (*Point*, *Multipoint*, *Line*, *Multiline*, *Polygon*, *Multipolygon*).

A motivação para a integração de um *Data Warehouse* com Operações OLAP e SIG para Agricultura de Precisão é uma simples consulta visual dos dados, até a análise do comportamento dos recursos administrados para o melhor aproveitamento e rendimento da produção agrícola. A análise pode ser realizada por meio da inspeção gráfica, de forma fácil, intuitiva e dinâmica, podendo ser agregado o conhecimento intangível de especialistas com a análise detalhada dos dados para produzir informação e até mesmo conhecimento. No próximo capítulo será abordado as tecnologias e conceitos que pretendem viabilizar a integração proposta.

Tecnologias Abertas e Extensíveis Aplicadas ao Suporte a Decisão na Agricultura

3.1 Introdução

O capítulo anterior apresentou diversos conceitos envolvendo suporte a decisão em agricultura. A implementação, bem como a integração de um sistema de apoio a decisão em agricultura, por meio dos conceitos de agricultura de precisão, *Data Warehouse* e SIG não é trivial. Exige uma arquitetura de software que possa suportar desde as camadas de mais baixo nível como banco de dados, até o alto nível como interface do usuário.

O uso de tecnologias abertas e extensíveis tem o papel de permitir o reuso de soluções existentes, bem como integrar com as demais camadas, e por fim estender suas funcionalidades afim de atender os requisitos de cada projeto. Por consequência do uso de tecnologias abertas e extensíveis, o produto final, também compartilhará destas características, sendo possível ser utilizada em projetos futuros. Para isso buscou-se tecnologias que permitam a implementação dos conceitos abordados no capítulo anterior, bem como sejam abertas e extensíveis:

- Ferramenta para processamento ETL (*Extraction, Transform and Load*) Geokettle (BARD, 2009);
- Banco de dados PostgreSQL com a extensão Postgis (OSGEO, 2016);

- Servidor OLAP Geomondrian (SPATIALYTICS, 2016);
- Biblioteca para consultas MDX OLAP4J (HYDE, 2016);
- API JSON para consulta Mandoline (ROBERT, 2014a);
- Aplicação web Geonode (Open Source Geospatial Foundation, 2016);
- API para consulta multidimensional (ROBERT, 2014d);
- Servidor de filtro (ROBERT, 2014c);
- Biblioteca para manipulação de documentos baseados em dados D3.js (DC.js, 2016);
- E por fim a biblioteca DC.js para gerar gráficos (DC.js, 2016).

3.2 Etapas ETL

O processo de extração, transformação e carregamento (ETL) permite a importação de dados não modelados em bases de dados operacionais, para uma base modelada multidimensionalmente de um *Data Warehouse*. As etapas ETL podem ser observadas na Figura 10 (MCGUIRE *et al.*, 2008).

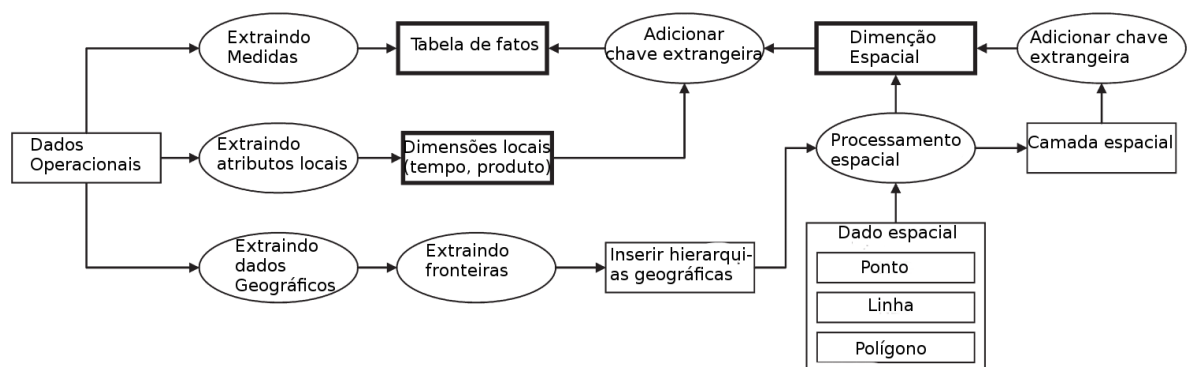


Figura 10: Modelo de fluxo de trabalho geral para a criação de um Data Warehouse espacial

Fonte: Adaptado de McGuire *et al.* (2008)

A primeira etapa é composta pelo conjunto de dados operacionais, originados por sensores aplicados na colhedora, como visto na Figura 4 da Seção 2.3. Os dados operacionais não modelados são apresentados em uma única tabela, com os componentes: geográfico (ponto, linha ou polígono), data e hora, variedade, cultura, operador e máquina. O seguinte passo é extrair os identificadores de cada medida, e agrupar os dados de acordo com cada dimensão modelada. Após os dados sumarizados pelas dimensões modeladas, inicia-se o passo para carregar os valores de cada dimensão e por fim carregar as medidas da tabela de fatos, com as respectivas chaves de suas dimensões.

A dimensão espacial exige um cuidado especial, afim de tratar corretamente os processos de agregação de dados, obter as fronteiras dos talhões e fazendas, para finalmente carregar os dados da hierarquia (talhão, fazenda, região rural).

Para a realização das operações de ETL foi utilizado a ferramenta GeoKettle (BADARD, 2009). Esta ferramenta é orientada a metadados dedicada à integração de diferentes fontes de dados espaciais ou não, para a construção ou atualização de *Data Warehouses*. A ferramenta suporta diversas fontes de dados, como arquivos csv (*Comma-separated values*), arquivos SHAPEFILE (Environmental Systems Research Institute, Inc, 1998), XML, e outros formatos de texto, e também oferece suporte para entrada de dados de diferentes SGBD (Sistemas Gerenciadores de Banco de Dados). Também suporta diversas transformações, como seleção de valores, adição de constantes, sumarização por agrupamento, cálculos e funções para objetos espaciais (união, interseção, bordas). E finalizando o processo, permite a saída de dados em arquivos de texto, csv, xml, tabela de banco de dados.

3.3 PostgreSQL e PostGIS

O PostgreSQL é um SGBD open source, sob licença BSD¹, inicialmente desenvolvido pela Universidade da Califórnia (Berkeley), com suporte para transações, chaves estrangeiras, vistas, gatilhos (triggers) e procedimentos.

O suporte ao armazenamento de dados geográficos é dado pela extensão *PostGIS*, que permite armazenamento, operações espaciais e criação de índices. O *PostGIS* é distribuído sob a licença GNU GPL (General Public Licence)², segue as certificações SFS (Simple Feature Specifications) do OGC e fornece funções de geoprocessamento através de comandos SQL.

O *PostgreSQL* juntamente com o *PostGIS* permite armazenar e manipular pontos, linhas, polígonos, multipontos, multilinhas e multi-polígonos por meio de comandos SQL. Assim devido as características citadas o SGBD *PostgreSQL/PostGIS* foi utilizado para compor a arquitetura e prover as funções de base de dados.

¹Licença BSD é uma licença de código aberto inicialmente utilizada nos sistemas operacionais do tipo Berkeley Software Distribution (um sistema derivado do Unix); o texto da licença é considerado como de domínio público e pode ser modificado sem nenhuma restrição, permitindo que o software distribuído sob esta licença seja incorporado em produtos proprietários.

²GNU GPL é a designação da licença para software livre idealizada por Richard Stallman no final da década de 80 do século XX, na âmbito do projecto GNU da Free Software Foundation (FSF).

3.4 Geomondrian

Geomondrian (SPATIALYTICS, 2016) é um servidor OLAP espacial de código aberto que implementa as funcionalidades do servidor OLAP *Mondrian* (HYDE, 2016) e estende as funções espaciais, como: (i) tipo de dado *Geometry* (OGC, 2007), que é utilizado para definir geometrias (ponto, linha e polígono); (ii) extensão da linguagem MDX para suportar operações espaciais e extensão do esquema XML de representação de metadados do *Mondrian* para suportar o tipo de dado *geometry*.

Com o servidor *Geomondrian* as consultas analíticas podem ser construídas abordando o aspecto espacial, através de operações de distância, interseção, disjunção, centro de um polígono, entre outras. Os metadados do cubo espacial é armazenado em formato XML e permite a manipulação e criação de forma textual ou por ferramentas compatíveis, como por exemplo *Mondrian Schema Workbench* distribuído sob a licença EPL (*Eclipse Public License*)³. O servidor *Geomondrian* é compatível com o SGBD espacial *PostgreSQL/PostGIS*.

3.5 OLAP4J e Mandoline

OLAP4J é uma API para realizar conexão no servidor SOLAP *Geomondrian*. Esta API permite o desenvolvimento de aplicações analíticas de forma facilitada e rápida, devido a sua interface de programação.

Esta API estende a conexão com o banco de dados JDBC⁴. Por exemplo, a classe *OlapConnection* estende a classe *java.sql.Connection*, a partir do qual pode-se criar declarações *OlapStatement* e executar uma consulta que irá resultar em um *CellSet* (semelhante ao *java.sql.ResultSet*).

Através da API OLAP4J Robert (2014a) desenvolveu um servidor que simplifica as consultas multidimensionais por meio de troca de mensagens JSON sobre soquetes⁵. Este servidor provê a interface para a camada da aplicação WEB.

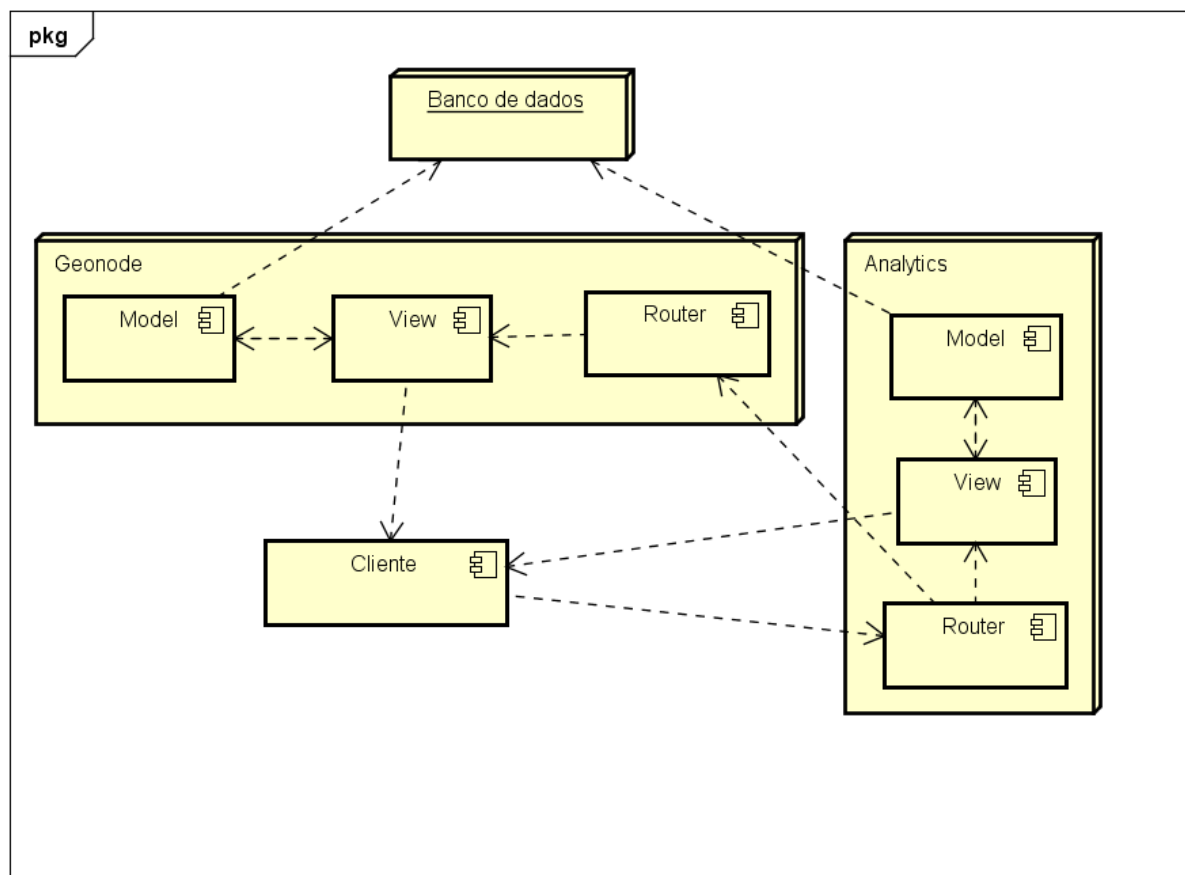
³A EPL foi criada pela Eclipse Foundation para substituir duas licenças anteriores à criação da fundação: a IBM Public Licence de 1999 que foi posteriormente substituída pela Common Public Licence.

⁴Java Database Connectivity ou JDBC é um conjunto de classes e interfaces (API) escritas em Java que fazem o envio de instruções SQL para qualquer banco de dados relaciona

⁵Soquete é um ponto final de um link de comunicação bidirecional entre dois programas em execução na rede. Um soquete está vinculado a um número de porta para que a camada TCP possa identificar o aplicativo para o qual os dados estão destinados.

3.6 Geonode e Analytics

O módulo *Analytics* estende as funcionalidades do software GeoNode (Open Source Geospatial Foundation, 2016). O GeoNode é uma aplicação e plataforma Web para desenvolver sistemas de informação geoespacial (SIG) e para implementar infra-estruturas de dados espaciais. A funcionalidade adicional proporcionada pelo módulo *Analytics* é o ambiente de análise, o qual proporciona a construção e manipulação analítica dos dados multidimensionais. A interface entre os módulos pode ser analisada na Figura 11.



powered by Astah

Figura 11: Diagrama de implementação da camada Aplicação WEB

Como mostrado na Figura 11 o módulo *Analytics* estende GeoNode e processa as solicitações e provê a interface, controle e modelos de dados para as funcionalidades analíticas. As demais solicitações são enviadas para o *GeoNode*.

3.7 API de consulta

Esta API auxilia consultas no cubo OLAP. Ela comunica-se com a camada de aplicação através de mensagens JSON sobre o protocolo http. A API implementa as seguintes funciona-

lidades para explorar o cubo do *Data Warehouse* fornecido pela interface do módulo *Analytics*(ROBERT, 2014d):

init : inicializa a API e dependências;

getSchemas : consulta todos os esquemas modelados no servidor de consultas OLAP Geomondrian;

getMesures : retorna as medidas de um dado cubo e esquema. As medidas são membros de apenas um nível e hierarquia da dimensão com o tipo medida mostrada

getCubesAndMeasures : retorna uma lista de cubos e suas respectivas medidas;

getDimensions : Especificado um determinado cubo, retorna as suas dimensões, com seus respectivos nomes e tipos (espacial, temporal, padrão)

getLevels : retorna uma lista de níveis de uma determinada hierarquia;

getMembers : retorna a lista de membros de uma dada dimensão:

Para manipular os dados para prover a funcionalidade de *drill-down* e *roll-up*, é armazenado o objeto em uma estrutura de pilha para cada nível exibido. Por exemplo, se no início da pilha conter o nó Brasil, e for realizado a operação de *drill-down* no Paraná, será adicionado o nó Paraná na pilha, assim sucessivamente. Os objetos empilhados contém as propriedades geométricas (como por exemplo o polígono que define os limites do estado do Paraná), e propriedades não geométricas (como por exemplo o número de habitantes no estado do Paraná). Desta forma é possível navegar entre os objetos, e sobre estes recuperar as informações ou até mesmo realizar outras operações como *slice* ou *dice*.

3.8 Crossfilter

A recuperação de dados é realizada por meio de consultas retornadas do servidor OLAP (Geomondrian). As operações de *drill-down*, *roll-up*, *slice*, são realizadas por meio dos parâmetros da consulta, sendo o retorno os dados consultados e armazenados no navegador (cliente). O esquema da Figura 12 descreve esse processo de recuperação.

Na Figura 12 os objetos da consulta são armazenados e processados pelo lado do cliente, por meio do módulo do Crossfilter. Para o processamento dos gráficos é utilizado as bibliotecas DC.js (DC.js, 2016) e D3.js (BOSTOCK, 2016).

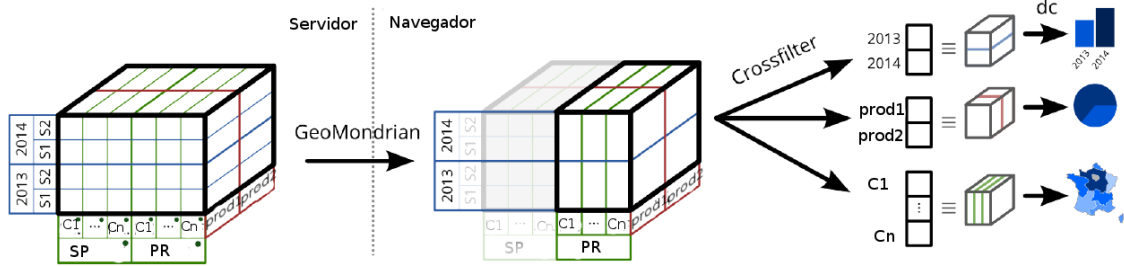


Figura 12: Solução para realização dos filtros e agregações via consulta
Adaptado de Robert (2014c)

Devido ao fato do processamento, assim como os dados estarem no lado do cliente, o tempo de resposta para um pequeno cubo de dados é rápido (< 10 ms). Porém com o aumento do número de membros de cada dimensão, o número de linhas tende a crescer exponencialmente, por exemplo, se temos a dimensão tempo com 10 anos, a dimensão espaço com 20 cidades e a dimensão produto com 50 produtos temos 10.000 linhas a serem processadas e sumarizadas para a construção dos gráficos. Caso a dimensão tempo seja alterada de anos para meses do ano, teremos $(120 \times 20 \times 50)$ 120.000 linhas. Portanto, esta solução é indicada para para conjunto de dados "pequenos", permitindo uma maior iteratividade e agilidade para processar e analisar informações.

3.9 Crossfilter server

Para um conjunto de dados maiores, pode ser otimizado removendo o cubo de dados intermediário, onde o módulo Crossfilter calculará diretamente a projeção para cada dimensão. Para manter a compatibilidade com as dependências e especificações do *Crossfilter*, bem como as bibliotecas DC.js e D3.js, a solução implementa um servidor Crossfilter, utilizando a API de consulta ao Geomondrian para calcular as projeções e agregações, como pode ser observado na Figura 13. Ainda, toda a carga de processamento dos dados é realizada no lado do servidor, cabendo ao lado do cliente apenas a construção dos gráficos, mapas e tabelas.

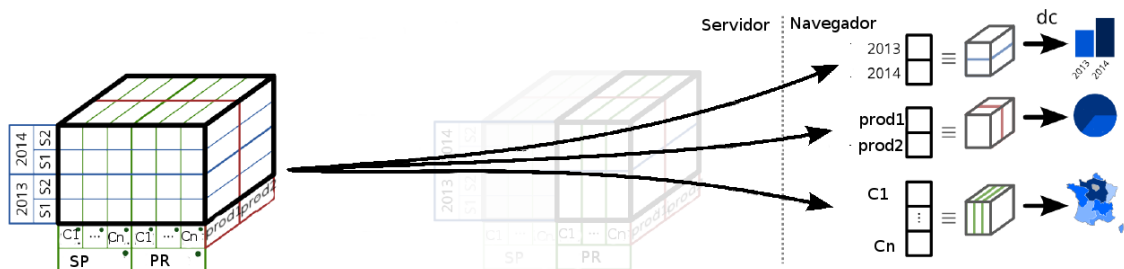


Figura 13: Solução para realização dos filtros e agregações via servidor
Fonte: Adaptado de Robert (2014c)

A desvantagem desta solução, é a necessidade de recarregar os dados a cada vez que for realizado uma filtragem, por exemplo ao ser mostrado um cubo com as dimensões tempo (anos), espaço (cidades) e produto, ao realizar uma operação de drill-down no ano de 2010 por exemplo, será necessário recarregar um novo cubo. Portanto, esta solução tem maior eficácia em cubos com maiores volumes de dados, porém haverá uma deficiência em processar operações OLAP em tempo de execução.

3.10 DC.js e D3.js

DC.js é uma biblioteca Javascript que oferece suporte nativo ao Crossfilter, permitindo a exploração eficiente em conjunto de dados multi-dimensionais. Já a biblioteca D3.js é utilizada em conjunto com a DC.js para renderizar gráficos em formato SVG⁶ e com estilização CSS⁷. Os gráficos podem ser implementados de forma iterativa, por exemplo, pode-se adicionar o evento de click em qualquer elemento do gráfico para realizar o filtro pelo item clicado.

3.11 Considerações

Neste capítulo foi analisado um portfólio de tecnologias abertas e extensíveis, que podem ser utilizadas para compor uma arquitetura de software, capaz de implementar os conceitos abordados no Capítulo 2 (Sistemas de Suporte à Decisão em Agricultura). Apesar destas tecnologias oferecerem os recursos necessários, ainda se faz necessário a implementação e integração destas, bem como a sua aplicação no contexto de dados agrícolas.

No próximo capítulo será analisado o contexto de sistemas de integração DW, AP, e SIG, definindo os requisitos para o suporte a decisão em agricultura. Também será apresentado trabalhos correlatos e sua análise de acordo com os requisitos levantados na contextualização.

⁶SVG é a abreviação de Scalable Vector Graphics, ou em português gráficos vetoriais escaláveis. É uma linguagem XML para descrever de forma vetorial desenhos e gráficos bidimensionais estático, dinâmico ou animado.

⁷Cascading Style Sheets (CSS) é uma linguagem de folhas de estilo utilizada para definir a apresentação de documentos escritos em uma linguagem de marcação, como HTML ou XML.

Requisitos de Integração e Trabalhos Correlatos

4.1 Introdução

Os conceitos apresentados abordando *Data Warehouse* e sistema de informação geográfica no escopo da agricultura de precisão resultou nos seguintes requisitos: Camada de apresentação, Modelagem de dados agrícolas, Modelagem espacial, Modelagem multidimensional, Integração com Agricultura de Precisão e Linguagem de consulta. A seguir será detalhado cada um dos requisitos.

4.2 Requisitos de Integração

A camada de apresentação é a interface com a qual o gerenciador da área agrícola poderá realizar as operações que fornecerão subsídio para auxiliar a tomada de decisão. A camada de apresentação deve seguir os requisitos não funcionais levantados no tópico sistemas agrícolas para agricultura de precisão.

A interface deve atender as necessidades de um gerente agrícola, de forma simples e funcional, permitindo o total controle dos parâmetros e funções de uma análise de cenário. A tomada de decisão apresenta-se como um processo complexo e difícil estabelecimento de padrões. Portanto, sistemas que pretendem dar apoio à tomada de decisão devem ser visualmente intuitivos, para que os gerentes agrícolas concentrem seus esforços no seu campo específico de

atuação. Assim, o software tende a ter uma curva de aprendizagem rápida e contribui com o trabalho do especialista.

Outro ponto levantado na camada de apresentação é a flexibilidade. O processo de construção de um caminho de análise de cenário apresenta variação de cenário para cenário. Ainda com relação à flexibilidade, especialistas apresentam um conhecimento advindo de experiências e práticas intangíveis por parte do software, mas que este deve permitir o uso e inclusão de conhecimento durante o processo de análise e geração de novos conhecimentos.

A modelagem de dados agrícolas deve ser realizada com base no negócio agrícola, levando em conta todo o processo monitorado e coleta de dados. A manipulação de forma correta dos dados depende da modelagem e interpretação dos mesmos. Os dados podem ser originados de varias fontes, como uma colhedora, um trator, um implemento de pulverização, entre outros. Também, podem apresentar formatos diferentes, por exemplo, unidades de conversão.

O processo de modelagem deve abranger todo o ciclo de produção da área agrícola, que envolve a preparação do solo, o plantio ou semeadura, fertilização, controle de pragas e ervas daninhas, e finalizando com a colheita.

Todas estas operações realizadas na área agrícola são armazenadas em uma base operacional, a qual é separada da base orientada a assunto, utilizada para o sistema de suporte a decisão. A modelagem espacial dos dados é requerida devido à natureza dos dados agrícolas, onde são identificados além dos dados (como a quantidade de grãos colhida) a posição geográfica dos mesmos. As condições geográficas devem ser representadas por meio dos recursos que o SIG oferece.

Um dado geográfico deve ser representado por um conjunto de objetos identificáveis e localizáveis. Os objetos podem ser classificados como pontos, curvas, superfícies ou sólidos. Assim, a composição destes elementos pode formar o objeto necessário para representar as informações agrícolas.

A modelagem multidimensional é baseada em assunto, diferentemente da modelagem operacional dos dados agrícolas que é relacional. Devido a isso, os dados agrícolas necessitam passar por um processo de ETL para incorporação na base de dados multidimensional. A extração consiste em recuperar os dados das bases operacionais. Já a transformação é encarregada de realizar os ajustes, conversões e agregações necessárias para a adequação no modelo multidimensional.

As dimensões encontradas na atividade agrícola podem ser divididas como: espaço (talhão, gleba, etc.), consistindo na localização geográfica; dimensão produto, representando as

informações da cultura; e a dimensão tempo, correspondendo a safra (ano, época).

Cada dimensão pode conter um ou mais atributos que individualizam uma representação única da dimensão, como por exemplo, a dimensão produto contém os atributos de cultura, variedade, etc. Já a informação é chamada de fato. Um fato deve ser identificado por um conjunto de valores das dimensões. Por exemplo, o fato produtividade possui um tempo (safra), um espaço (talhão) e um produto (cultura). Pelas manipulações e operações OLAP é possível analisar os fatos nas dimensões desejadas, e proporcionar visões de agregações, junções e subconjuntos, de forma a moldar os dados e produzir informação, assim como conhecimento.

A integração com Agricultura de precisão consiste na aplicação das soluções analíticas, auxiliando o gerenciamento de aplicações, manejos e operações no campo agrícola. É importante observar a variabilidade espacial que a agricultura de precisão tem como foco, denotada pela representação espacial, permitindo a manipulação dos dados agrícolas individualizada em cada ponto coletado.

O gerenciamento de recursos na agricultura de precisão é complexo. Por exemplo, como decidir a aplicação de adubo em áreas com menor desenvolvimento (consequentemente menor potencial produtivo), ou em áreas com um melhor desenvolvimento, mas que podem apresentar uma melhor resposta ao insumo aplicado. A resposta para essas questões varia em cada cenário, e para isso a ferramenta de análise multidimensional pode auxiliar na busca do conhecimento, para aplicar as taxas variáveis, para obtenção da melhor resposta, evitando desperdícios.

Finalmente, a linguagem de consulta é um requisito para possibilitar a flexibilidade e dinamismo do usuário para com a ferramenta analítica. Uma linguagem de consulta permite a execução de operações complexas, assim como a agregação do conhecimento advindo de experiências práticas não representado em dados, mas que o gerente ou especialista precisa incorporar.

Neste capítulo foi discutido os requisitos necessários para atender os conceitos de *Data Warehouse*, SIG e agricultura. No próximo capítulo serão analisados trabalhos correlatos a *Data Warehouse* espacial e/ou agricultura.

4.3 Trabalhos Correlatos à *Data Warehouse* espacial e agricultura

A integração da prática da AP com uma ferramenta para suporte a decisão em uma visão multidimensional, e o tratamento de dados espacialmente referenciados foi abordado por Vernier *et al.* (2013). Já outras literaturas, como Glorio *et al.* (2012), Silva *et al.* (2010) e Kyung,

Yom e Kim (2012)), abordam questões relacionadas a Ferramentas DW e OLAP, integrados com dados espaciais (SIG), os quais são relatados na sequência.

4.3.1 Personalização para um *Data Warehouse* espacial

A modelagem multidimensional, com dados espaciais proporciona a criação de uma estrutura para uso de técnicas de análise e manipulação de dados em sistemas de suporte a decisão. A proposta de Glorio *et al.* (2012) é a personalização, tornando a solução flexível ao contexto de utilização. A principal característica é o perfil para modelagem conceitual multidimensional espacial, seguindo para o contexto espacial, definindo características da modelagem espacial e uma linguagem para definição de regras de tomada de decisão, através da seleção do perfil de modelagem, adequado a cada cenário.

Para a descrição de diferentes cenários, é utilizado um estudo de caso, onde a modelagem para cada cenário exige diferentes representações espaciais. O primeiro caso aborda promoções de entrega, analisando a proximidade dos clientes da loja, ou seja, descrição espacial de endereços de cliente e de loja (*Point* cliente, *Point* loja e função *Distance()*). Outro cenário é a venda de produtos farmacêuticos, em relação a proximidade da cadeia de farmácias com hospitais, sendo necessária a localização da cadeia de farmácias e de hospitais (*Layer* hospitais, *Point* farmácia e função *Distance()*). Finalmente, o suporte à decisão em um Departamento de Recursos Humanos, para análise da produtividade dos funcionários, com relação ao transporte da casa para o local de trabalho. Para este cenário é utilizada a descrição espacial da casa dos funcionários, assim como as estações de metrô e ônibus, e a localização do local de trabalho (*Layer* estacao_metro, *Layer* estacao_onibus, *Point* local_trabalho, *Point* casa_funcionario e função *Distance()*). A personalização de cada cenário é realizada no processo da modelagem multidimensional, utilizando uma Linguagem PRML - *Personalization Rules Modeling Language*.

O trabalho de Glorio *et al.* (2012) possui como vantagem a modelagem flexível, possibilitando a construção de vários cenários. O foco deste trabalho não contempla agricultura, portanto não apresenta os requisitos relacionados a este escopo levantados neste capítulo. O próximo tópico abordará um trabalho relacionado à um *Data Warehouse* e as respectivas operações OLAP espaciais

4.3.2 Spatial Data Warehouse e Spatial OLAP

Spatial DW - SDW e *Spatial OLAP* - SOLAP (KYUNG; YOM; KIM, 2012) são ferramentas de monitoramento de base de dados geoespacial, compostas por tipos de dados vetor e

raster. Os autores apresentam a modelagem de uma base de dados multidimensional e critérios de atualização de dados geoespaciais para auxiliar o processo de tomada de decisão.

A modelagem foi realizada considerando o custo da atualização e publicação dos dados espaciais. É um fator crítico para tomada de decisão, na atualização e no processo de decisão, como a extensão das ruas, utilização das construções, uso do espaço e indicação de mudanças topográficas. A tabela da dimensão espacial foi construída, dividindo a localização em uma hierarquia de 6 (seis) níveis, em bairros até o nível de país. A dimensão estrutura, para os tipos de dados, foi dividida em 5 (cinco) níveis, de acordo com o nível de abstração. Como exemplos, têm-se ortoimagem, georeferenciado e *raster*. A dimensão de uso foi classificada em residência, comercial, industrial, educacional e outros.

A dimensão do uso do espaço foi classificada como áreas restrita, planejada e em uso. A dimensão temporal foi descrita como dos anos 2000 a 2009. A tabela de fatos incorpora os valores mensurados, para extensão das ruas, utilização das construções e utilização do espaço.

O aumento de dimensões ou no nível de hierarquia causa um aumento significativo no número de registros no DW, e no momento da modelagem deve-se verificar a disponibilidade dos dados para as dimensões, para evitar valores nulos, com consequente perda de credibilidade.

As vantagens do trabalho de Kyung, Yom e Kim (2012) são a camada de apresentação e o uso de *raster* e vetor. Não há apresentação de um linguagem de consulta. O enfoque deste trabalho também não contempla a agricultura. O próximo tópico aborda um trabalho de um *Data Warehouse* geográfico.

4.3.3 Geographical Data Warehouses

Silva *et al.* (2010) propôs um metamodelo para a descrição de DW geográficos, para integrar operadores multidimensionais e espaciais, com sintaxe única. Foi criado um conjunto de operadores especiais e uma interface gráfica que permite ao usuário visualizar o resultado das consultas. O modelo possui um esquema, e esse possui uma ou mais tabelas. Cada tabela é composta pela sua chave primária, colunas (atributos, medidas, medidas espaciais) e chaves estrangeiras. A dimensão pode ser classificada como convencional ou geográfica, onde a geográfica pode ser primitiva, como pontos, linhas polígonos, ou composta.

É definido todos os domínios de atributos e relações, como, domínio convencional e geográfico, atributos comuns ou localização ou geométricos, esquemas de dimensão do DW, das dimensões, geográfica, convencional, híbrida, esquema fato do DW, relação do fato do DW, esquema do DW geográfico, instância do DW geográfico, níveis dos membros, hierarquia de

dimensões, projeção de fatos, funções de agregação e cubo de dados.

Também, foi definida a linguagem de consulta multidimensional geográfica. A linguagem sobrepõe os operadores definidos no Padrão *MultiDimensional eXpressions* (MDX), sendo: *DrilldownLevel* (especialização de nível), *DrilldownMember* (especialização de membro), *DrillupLevel* (generalização de nível), *DrillupMember* (generalização de membro), *Ancestor* (antecessor), *Descendants* (descendentes), *Ascendants* (ascendentes), *Members* (membros do nível), *Children* (filho do membro), *Siblings* (irmãos do membro, incluindo o mesmo) e *All* (todos os níveis superiores da hierarquia).

Os operadores baseados no OGC são definidos como: *Distance* (distância entre dois membros), *Positional operators* (a posição em relação a outro membro), *Topological operators* (relação topológica), *Intersection* (interseção entre dois membros), *Union* (união entre dois membros), *Buffer* (todos os pontos cuja a distância desta geometria seja menor ou igual a uma distância especificada), *Area* (área do membro), *Length* (comprimento do membro) e *Equals* (igualdade entre membros).

O conjunto de operadores especiais corresponde a: *DrillOut* (para um dado nível, retorna os vizinhos de um membro), *HighestDistance* (Classifica os membros que estão dentro de uma dada distância em ordem decrescente), *LowestDistance* (Classifica os membros que estão dentro de uma dada distância em ordem crescente), *RankArea* (classifica pela área), *RankLength* (classifica pelo comprimento) e *Point* (todos os pontos centrais de um conjunto de polígonos).

O sistema apresenta uma solução de integração e normalização de dados geométricos, com uma linguagem de consulta com uma sintaxe integrada e concisa e a visualização de tabelas, mapas e gráficos.

As vantagens do trabalho de Silva *et al.* (2010) envolvem a camada de apresentação, a definição de um metamodelo e uma linguagem de consulta multidimensional geográfica. O próximo trabalho correlato a ser apresentado é o único levantado por esta pesquisa que aborda questões agrícolas.

4.3.4 EIS Pesticidas

A motivação de Vernier *et al.* (2013) foi a redução do uso de pesticidas através do entendimento das interações da água, do uso do solo, e do ambiente. Para isso, foi criado um sistema de informação ambiental (*EIS – Environmental Information System*) para pesticida. O sistema permite a análise das atividades agrícolas, além das características das bacias hidrográficas. Para a caracterização das práticas foi utilizado objetos espaciais específicos. A aná-

lise permitiu encontrar resultados em diferentes níveis de interação, para diferentes dimensões. Como exemplo, tempo, área de superfície pulverizada ou tipo de pesticida.

O sistema foi modelado na Linguagem UML, com ferramentas espaciais adaptadas. Para o conceito espacial foi usado um delineador espacial predefinido, considerando atividades homogêneas para rotação, operações agrícolas e sequência de gerenciamento de campo. As dimensões modeladas foram: aplicação de pesticida, fórmula comercial, temporal, área agrícola, e a composição da bacia hidrográfica. Definidas as dimensões, o cubo foi construído.

Segundo Vernier *et al.* (2013), as ferramentas de DW e espaciais permitiram a manipulação simultânea de vários indicadores que descrevem o ambiente agrícola. O fator temporal permite avaliar a evolução do sistema. E, também, uma forma efetiva de avaliar os impactos da agricultura na bacia hidrográfica, por meio de avaliação da intensidade de aplicação de pesticida em um determinado período.

As vantagens do trabalho de Vernier *et al.* (2013) são a definição de uma camada de apresentação e apresentar uma modelagem de dados agrícolas. A modelagem de dados espaciais não atende a operações espaciais mais complexas, não há uma linguagem de consulta.

4.4 Abordagem Comparativa

Segundo Fountas *et al.* (2015)), a agricultura gera grande quantidade de dados, para ser utilizado de forma eficaz, no manejo da cultura, em um local específico. Assim, a tendência em desenvolvimento de aplicações e arquiteturas mais complexas na agricultura de precisão, envolvendo gerenciamento espacial e temporal. Portanto, há necessidade da agricultura de precisão gerenciar seus dados e informações, apresentando-os de forma intrínseca a característica espacial, gerando conhecimento e dando suporte a tomadas de decisão em nível estratégico.

As soluções encontradas constituem-se boas propostas em seus respectivos escopos, mas não atende a agricultura de precisão, fornecendo ferramenta de suporte a decisão efetiva e gerenciamento agrícola (Quadro 4.1).

4.5 Considerações

Neste capítulo foi apresentado os requisitos para um sistema de suporte a decisão em Agricultura, assim como pesquisas que possuem pontos em comum com a presente dissertação. Considerando que a agricultura gera grande quantidade de dados, a tendência em desenvolvimento de aplicações e arquiteturas mais complexas na agricultura de precisão, envolve

Quadro 4.1: Análise das propostas de integração DW, SIG e AP

	PSDW	SDW/SOLAP	GDW	EIS Pesticides
Camada de apresentação	Não	Sim	Sim	Sim
Modelagem de dados agrícolas	Não	Não	Não	Sim
Modelagem espacial	Sim	Sim	Sim	Não
Modelagem Multidimensional	Sim	Sim	Sim	Sim
Integração com AP	Não	Não	Não	Não
Linguagem de consulta	Não	Sim	Sim	Não

Fonte: O autor

gerenciamento espacial e temporal. Portanto, há necessidade de a agricultura de precisão gerenciar seus dados e informações, apresentando-os de forma intrínseca à característica espacial, gerando conhecimento e dando suporte a tomadas de decisão em nível estratégico.

Diante da necessidade de gerenciamento espacial e temporal, em nível estratégico na agricultura, foram especificados os requisitos necessários para análise multidimensional e espacial na agricultura. Também, foram levantadas as metodologias existentes, utilizadas para este fim. Contudo, foram encontradas soluções parciais nos quesitos de camada de apresentação, modelagem de dados agrícolas, modelagem espacial, modelagem multidimensional, integração com agricultura de precisão e linguagem para consulta. Assim, constatou-se a necessidade de uma metodologia para promover o suporte a tomada de decisão, no meio agrícola

No próximo capítulo será apresentado os métodos utilizados para a integração das operações analíticas e geográficas e uma interface gráfica que permite realizar as operações SOLAP. Todos padrões apresentados no próximo capítulo são abertos, satisfazendo os objetivos de padronização, extensibilidade e independência de plataforma.

Material e Métodos

5.1 Introdução

Ao longo do desenvolvimento desta pesquisa foram constatados os pontos de interseção entre as práticas da agricultura de precisão, com o suporte a decisão por meio de um data warehouse e as características espaciais promovidas pelos sistemas de informação geográfica.

O método proposto é baseado em tecnologias abertas e extensíveis, sendo elas: banco de dados PostgreSQL com a extensão Postgis (OSGEO, 2016), servidor OLAP Geomondrian (SPATIALYTICS, 2016), biblioteca para consultas MDX OLAP4J (HYDE; KLAUANS, 2011), API JSON para consulta Mandoline (ROBERT, 2014a), Aplicação web estendida do Geonode (Open Source Geospatial Foundation, 2016), API para consulta multidimensional (ROBERT, 2014d), servidor de filtro (ROBERT, 2014c), biblioteca para manipulação de documentos baseado em dados D3.js (BOSTOCK, 2016) e por fim a biblioteca DC.js (DC.js, 2016) para gerar gráficos.

5.2 Métodos

Este método foi elaborado baseado nas tecnologias abordadas no Capítulo 3 (Tecnologias Abertas e Extensíveis Aplicadas ao Suporte a Decisão na Agricultura), para implementar os recursos necessários para integrar a agricultura de precisão com sistema de apoio a decisão baseado em *Data Warehouse*, *OLAP* e *SIG*. O método pode ser analisado na Figura 14.

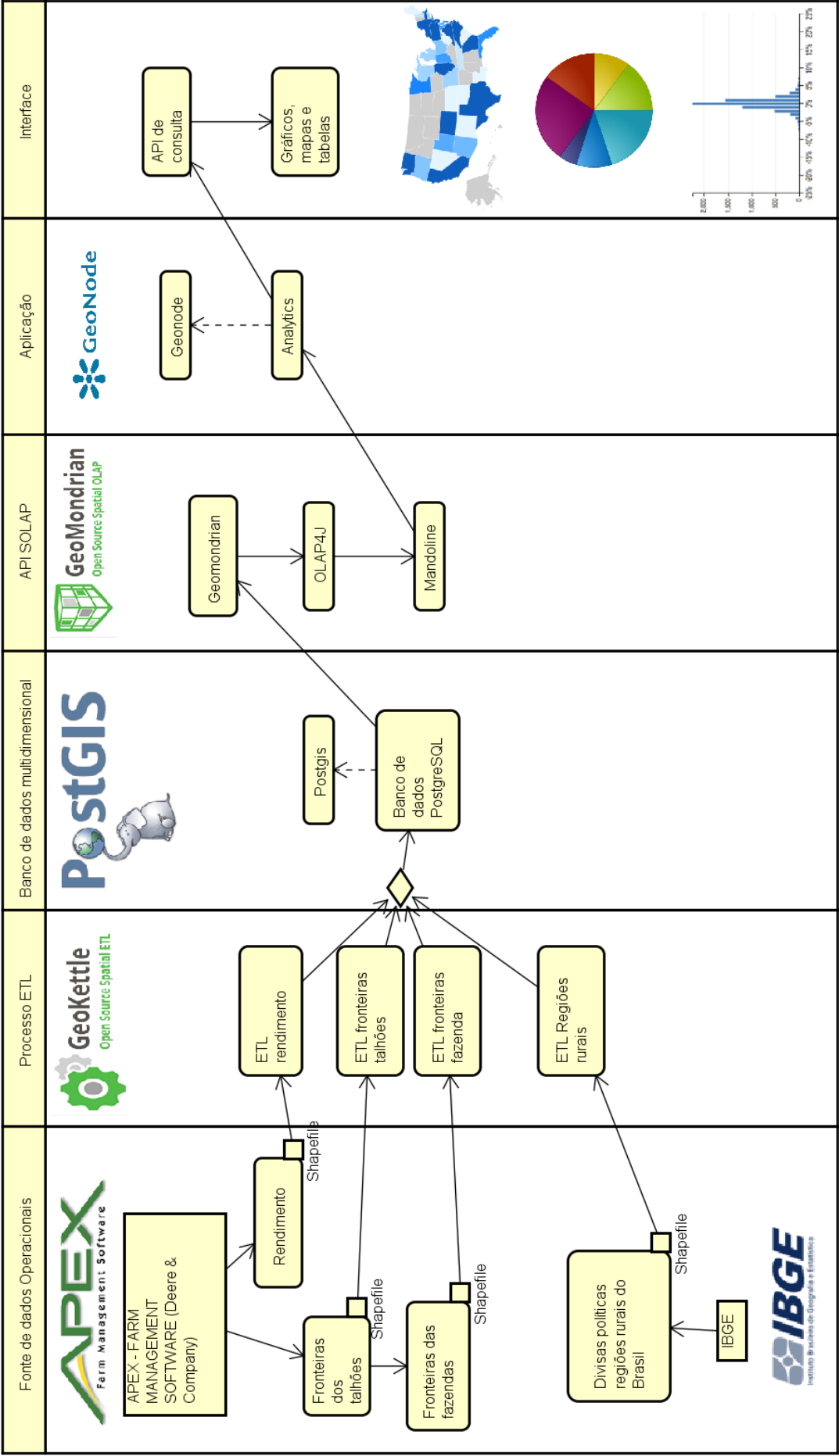


Figura 14: Método proposto baseado em tecnologias abertas e extensíveis para integração de um ambiente *Data Warehouse*, *OLAP* e SIG para agricultura

Pode-se observar na Figura 14 que o processo é constituído de seis principais módulos, sendo eles: fonte de dados operacionais, processo ETL, Banco de dados multidimensional, API SOLAP, Aplicação e Interface. Todas essas ferramentas e tecnologias foram instaladas e configuradas em uma máquina com processador core I5, 1TB de HD, 8GB de RAM e sistema operacional Linux Ubuntu 14.04 LTS.

5.2.1 Fonte de dados operacionais

A propriedade de onde foram obtidos os dados para o estudo de caso é localizada no estado do Paraná, município de Pirai do sul, nas coordenadas 24°20'46"S de latitude e 50°08'06"W de longitude e altitude média de 950 metros acima do nível do mar. A propriedade está dividida em 3 fazendas, sendo a primeira com 28 talhões, a segunda 16 talhões e a terceira 4 talhões. Clima subtropical úmido mesotérmico, temperatura anual média entre 17 e 18°C e precipitação média anual de 1750 mm. As principais culturas cultivadas são a soja, milho e trigo. A fazenda utiliza um software proprietário para a gestão de operações agrícolas chamado APEX (Deere & Company, 2011), que permite a exportação para o formato SHAPEFILE (Environmental Systems Research Institute, Inc, 1998) utilizado para a importação dos dados no banco multidimensional analítico pelo processo ETL descrito na próxima seção.

Os dados referentes as regiões rurais do Brasil foram obtidos por meio do projeto do IBGE (Instituto Brasileiro de Geografia e Estatística) o qual definiu 104 Regiões Rurais Brasileiras (Instituto Brasileiro de Geografia e Estatística, 2015). Estes dados são disponibilizados em formato SHAPEFILE, utilizado para entrada no processo ETL.

5.2.2 Processos ETL

Esses processos foram definidos de acordo com o modelo multidimensional estrela, sendo composto pela tabela de fatos rendimento, dimensão espacial, temporal e de produto (variedade plantada).

O carregamento foi realizado a partir de 8 arquivos SHAPEFILE, sendo 7 deles exportados pelo software APEX, e um obtido IBGE. O Quadro 5.1 apresenta os campos que cada arquivo possui:

A tabela de fatos foi composta pelas variáveis do arquivo de entrada rendimento, massa seca, massa úmida, e as chaves estrangeiras para as dimensões tempo, espaço e produto. A dimensões de tempo foi extraída da mesma tabela, sendo sumarizada em ano, mês, semana e dia. Já a dimensão produto também é oriunda da mesma tabela e sumarizada em variedade e

Quadro 5.1: Arquivos da base de dados operacionais e seus respectivos campos

Nome do arquivo	Campos dos dados
Regiões_rurais.shp	id_regiao, nome_regiao, geometry
Fronteiras_fazenda_paiquere.shp	Farm_name, geometry
Fronteiras_fazenda_Santa_branca.shp	Farm_name, geometry
Fronteiras_fazenda_Santa_Cruz.shp	Farm_name, geometry
Talhao_Paiquere.shp	talhao, Farm_name, geometry
Talhao_Santa_Branca.shp	talhao, Farm_name, geometry
Talhao_Santa_Cruz.shp	talhao, Farm_name, geometry
Fato_rendimento.shp	data_hora, elevacao, largura, velocidade, altitude, latitude, longitude, rendimento, massa seca, massa umida, variedade, cultura

Fonte: O autor

cultura.

A dimensão de espaço foi utilizado o mesmo software APEX, para exportar as fronteiras, tanto das fazendas, como dos talhões. Para finalizar o último nível da hierarquia da dimensão espaço foi composto pelas regiões rurais do Brasil, cuja a origem é do IBGE.

A limpeza dos dados foi realizada apenas uma simples filtragem de dados com valores não nulos nos campos rendimento, massa seca, massa úmida e umidade. Todos os passos de ETL foi realizado pelo software Open Source Geokettle (BADARD, 2009).

5.2.3 Banco de dados Multidimensional

A carga de dados resultante do processo ETL deve ser suprida pelo SGBD PostgreSQL, juntamente com a extensão PostGIS (OSGEO, 2016). A instalação foi realizada de acordo com o manual OSGEO Project (2016), e adicionadas as extensões e funções para atributos geográficos. O banco de dados foi modelado utilizando a topologia estrela, como visto na Figura 3 na Seção 2.3.

5.2.4 API SOLAP

O objetivo desta API é prover uma interface de comunicação entre a aplicação e o SGBD. Essa interface deve ser baseada na linguagem de consulta MDX, abordada na seção 2.4. Este módulo é composto pelo servidor de consultas multidimensionais *geomondrian*, a API OLAP4J para a estender as funcionalidades do *geomondrian*, e a API *mandoline*, para prover um serviço web de consultas mdx.

A instalação do projeto *geomondrian* foi realizada seguindo as recomendações da documentação disponível em Spatialytics (2016). Bem como a API OLAP4J também foi instalada utilizando as recomendações disponíveis em sua documentação (HYDE; K LAWANS, 2011). E por fim instalação do serviço do mandoline pode deve ser realizada por meio do código disponibilizado no repositório do github¹

A comunicação entre esta API e a aplicação deve ser baseada em troca de mensagens no formato JSON sobre sockets.

5.2.5 Aplicação

A camada de aplicação foi construída estendendo o aplicação geonode².

Para a instalação deste módulo, é requerido a instalação da aplicação geonode, seguindo as recomendações da documentação disponível em Open Source Geospatial Foundation (2016). E posteriormente a instalação de sua extensão analytics, disponível no repositório do github.

5.2.6 Interface

O módulo de interface foi utilizado os plugins D3.js (BOSTOCK, 2016) e DC.js (DC.js, 2016) para renderizar e construir gráficos de pizza, linha, de barra, contagem de palavras, tabelas e mapas. A documentação e instruções de instalação constam em suas respectivas documentações.

5.3 Considerações

Neste capítulo foram apresentados os métodos utilizados para a realização desta pesquisa. Que permite a configuração de um ambiente que suporta dados agrícolas, bem como permite as operações analíticas e multidimensionais, além de manipular dados geométricos. Por meio dos métodos utilizados, foi possível definir uma arquitetura que tem por objetivo facilitar a construção do ambiente analítico proposto. Esta arquitetura é definida no próximo capítulo.

¹Projeto de repositório online para distribuição e compartilhamento de software livre.

²O GeoNode é uma aplicação Web para desenvolver sistemas de informação geográfica (SIG) e implementar infra-estruturas de dados espaciais.

Resultados e Discussões

Analisando os conceitos abordados pela temática proposta, e revisão da literatura, notou-se grande dificuldade em encontrar padrões e arquiteturas que suportem a integração de um ambiente *Data Warehouse*, com as características espaciais, permitindo operações OLAP e abordando dados agrícolas.

Para tanto propõe-se que a principal contribuição deste trabalho é a definição de uma arquitetura aberta e extensível, que atenda as necessidades analíticas agrícolas, e possa ser implementada em diversos cenários de dados como em análise ao longo dos anos do desenvolvimento de uma variedade, impactos na produtividade ocasionados por manejos de adubação, entre outros.

6.1 Arquitetura AGRIBI

Utilizando os conceitos abordados até o presente capítulo, definiu-se que a análise de dados necessita de uma perspectiva geográfica. Propõe-se a arquitetura da Figura 15 chamada AGRIBI (Agricultural Business Intelligence), que permite a construção de uma ferramenta para suporte a decisão, abordando os conceitos de *Data Warehouse*, SIG e com operações OLAP, dividida em quatro camadas: Banco de dados, API SOLAP, APlicação WEB e Interface WEB. A camada de banco de dados é responsável pela modelagem física multidimensional dos dados, considerando a natureza geográfica dos mesmos. Já a camada API SOLAP é um serviço

WEB (Web service) que se comunica com a camada de dados através do padrão de acesso a dados XMLA (Microsoft and Hyperion Solutions, 2002) sobre o protocolo HTTP (FIELDING *et al.*, 1999). A camada de Aplicação WEB é uma extensão da plataforma de desenvolvimento web para aplicações geoespaciais Geonode (Open Source Geospatial Foundation, 2016), que foi estendida para a implementação da função de análise BI (*Business Intelligence*), que se comunica com a camada API SOLAP via sockets por mensagens JSON (JavaScript Object Notation) (CROCKFORD, 2006). E por fim a ultima camada é a de Interface WEB que tem a função de comunicação com o usuário final por meio de recursos visuais como gráficos, mapas, tabelas, nuvem de palavras, entre outros. Esta interface comunica através do protocolo HTTP por mensagens JSON (CROCKFORD, 2006). As próximas seções detalham cada componente da arquitetura proposta.

6.1.1 Banco de dados

O banco de dados analítico foi modelado seguindo a topologia de estrela, como visto na Figura 3 da Seção 2.3. A Figura 16 mostra a modelagem pé de galinha.

Para carregar os dados nesta base de dados definida na Figura 16, foram utilizadas as seguintes etapas para realizar o processo ETL definido no Capítulo 5:

Carregar as regiões rurais do Brasil: Esta etapa é composta por apenas 3 passos, como mostrado na Figura 17.

O primeiro passo é a entrada do arquivo SHAPEFILE, o segundo é uma conversão (cast) para transformar o id, que devido a leitura do arquivo é interpretado como caractere, para inteiro. O último passo é a inserção ou atualização dos dados na tabela `dim_regioes`.

Carregar as fronteiras das fazendas Esta etapa utiliza 3 arquivos, sendo um para cada fazenda como mostrado na Figura 18.

Na Figura 18 nota-se que para cada arquivo de fazenda são realizados os seguintes passos: entrada do arquivo SHAPEFILE, ordenação dos dados como pré-requisito para o passo seguinte que é agrupar e sumarizar os dados. Após é necessário adicionar a chave estrangeira para região. Concluído esses passos é realizada a inserção ou atualização em na tabela `dim_fazenda`.

Carregar as fronteiras de talhões para cada fazenda: Os passos desta etapa, semelhantes à anterior, podem ser observados na Figura 19, onde há como entrada 3 arquivos (um para cada fazenda), contendo as informações de nome e dado geográfico. O segundo e terceiro passo foram necessários devido aos nomes de talhões se repetirem entre fazendas diferentes, por isso foi adicionado um prefixo e concatenado ao nome do talhão, para marcar

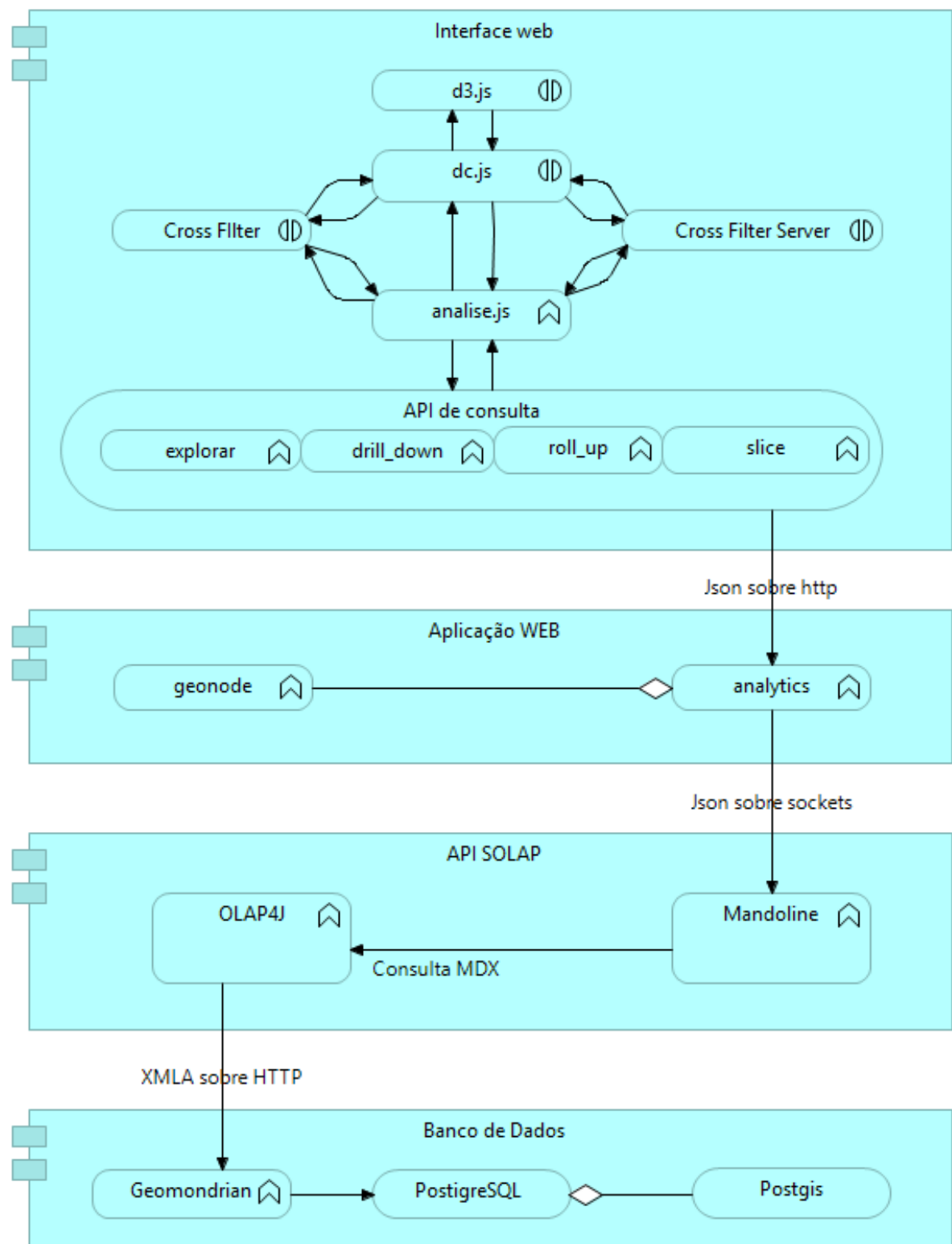


Figura 15: Arquitetura proposta para a integração de *Data Warehouse* e SIG em agricultura

de forma individual, e proporcionar a busca por nome. Já os passos seguintes são constituídos pela ordenação e agrupamento para sumarizar e finalmente inserir ou atualizar os dados na tabela *dim_talhao*.

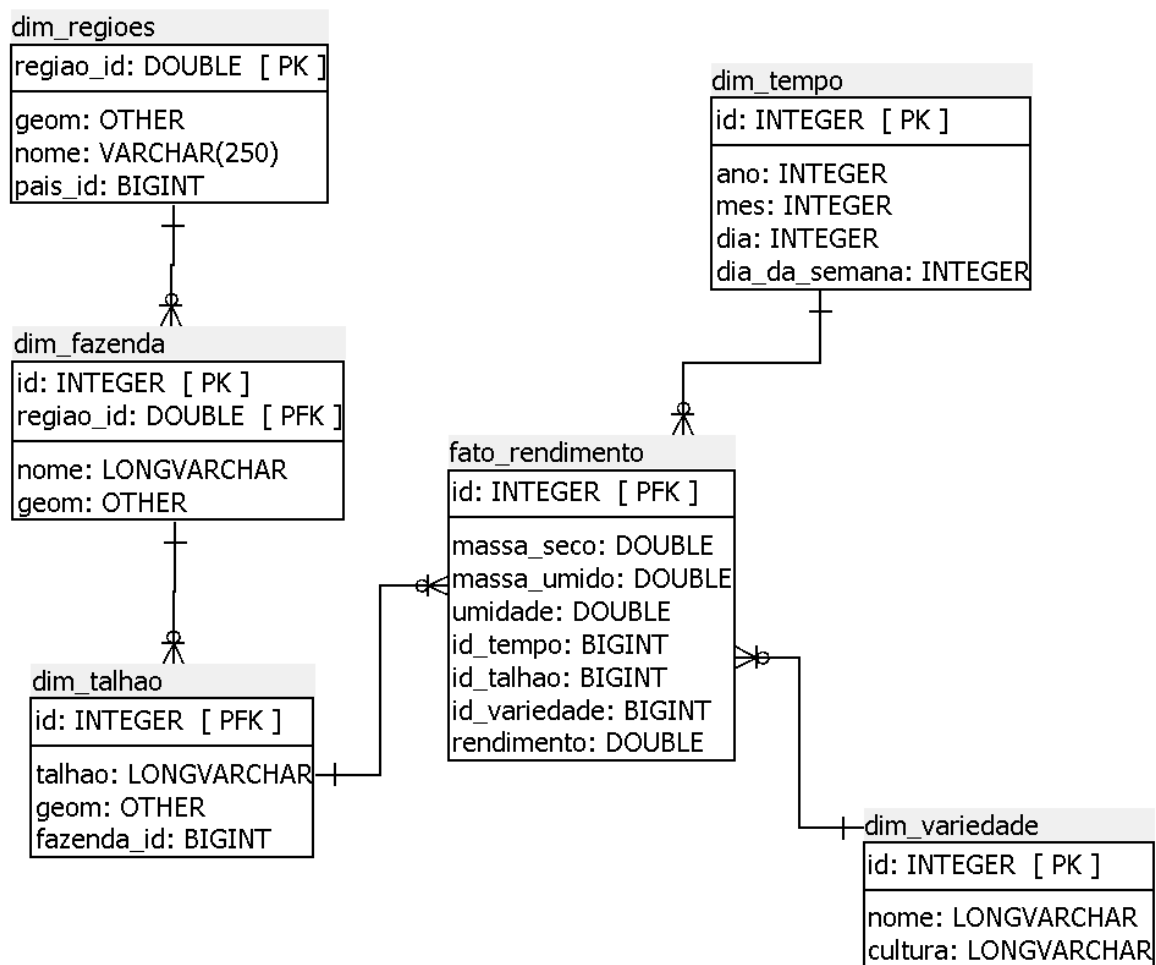


Figura 16: Modelagem estrela do banco de dados

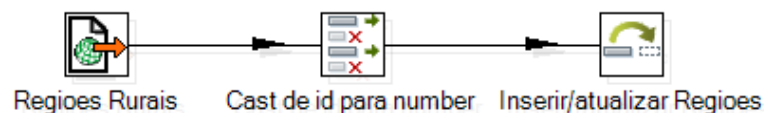


Figura 17: Processo de ETL modelado para inserir hierarquia regiões rurais na dimensão espacial

Carregar a tabela de fatos e as dimensões: A última etapa tem a tarefa de atualizar a tabela de fatos, bem como as tabelas das dimensões tempo e variedade, como pode-se observar na Figura 20, que contém os seguintes passos: carregar os dados do arquivo SHAPE-FILE, preparar o campo de data para o formato dd/MM/yyyy, calcular as variáveis ano, mês, dia da semana, e dia. O próximo passo é ordenar e agrupar os dados para sumarizar as variáveis de medidas. Após os dados sumarizados, realiza a inserção e atualização das dimensões de tempo (ano, mês, dia da semana e dia) e a dimensão variedade (variedade e cultura). O próximo passo deve aguardar que todos os itens das dimensões tempo e produto, estejam completas, por isso utilizou-se o passo chamado "Aguardar atuali-

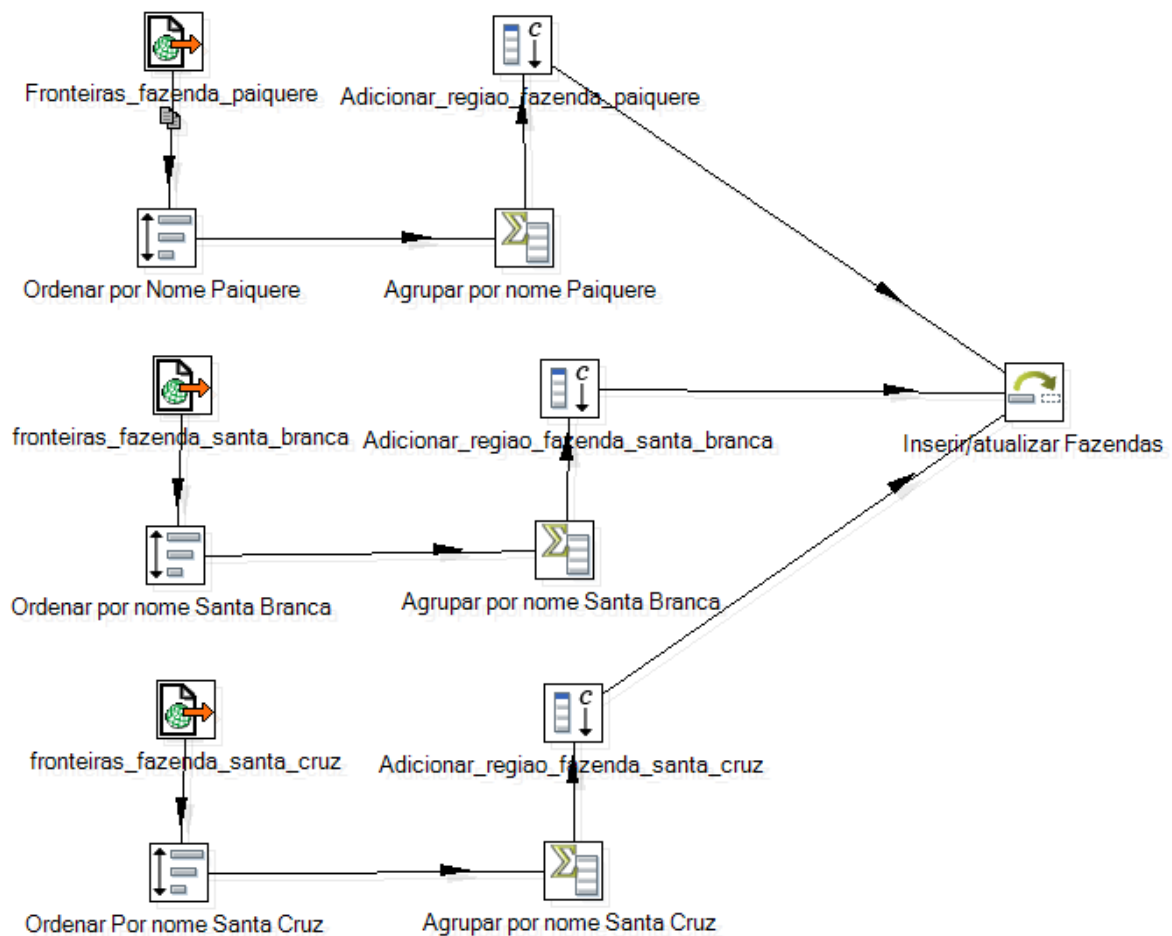


Figura 18: Processo de ETL modelado para inserir hierarquia fazenda na dimensão espacial

zação"(passo bloqueante). Após os passos de atualização serem concluídos, inicia-se os passo para buscar as chaves estrangeiras das dimensões variedade, tempo e talhão. Após encontrada estas chaves, o passo final é executado, atualizando ou inserindo os fatos na tabela fato_rendimento.

6.1.2 Aplicação

Como resultados da camada de aplicação pode-se destacar os componentes estendidos, afim de prover uma aplicação, integrada com o aplicativo *geonode*, assim implementando as funcionalidades analíticas principais:

- comportamento *drill-down/roll-up* entre os diferentes níveis do cubo OLAP aplicados nos itens de gráficos, tabelas e mapas;
- aplicar os filtros e redesenhar mapas, tabelas e gráficos
- mostrar uma legenda para a escala de cores usada em um gráfico

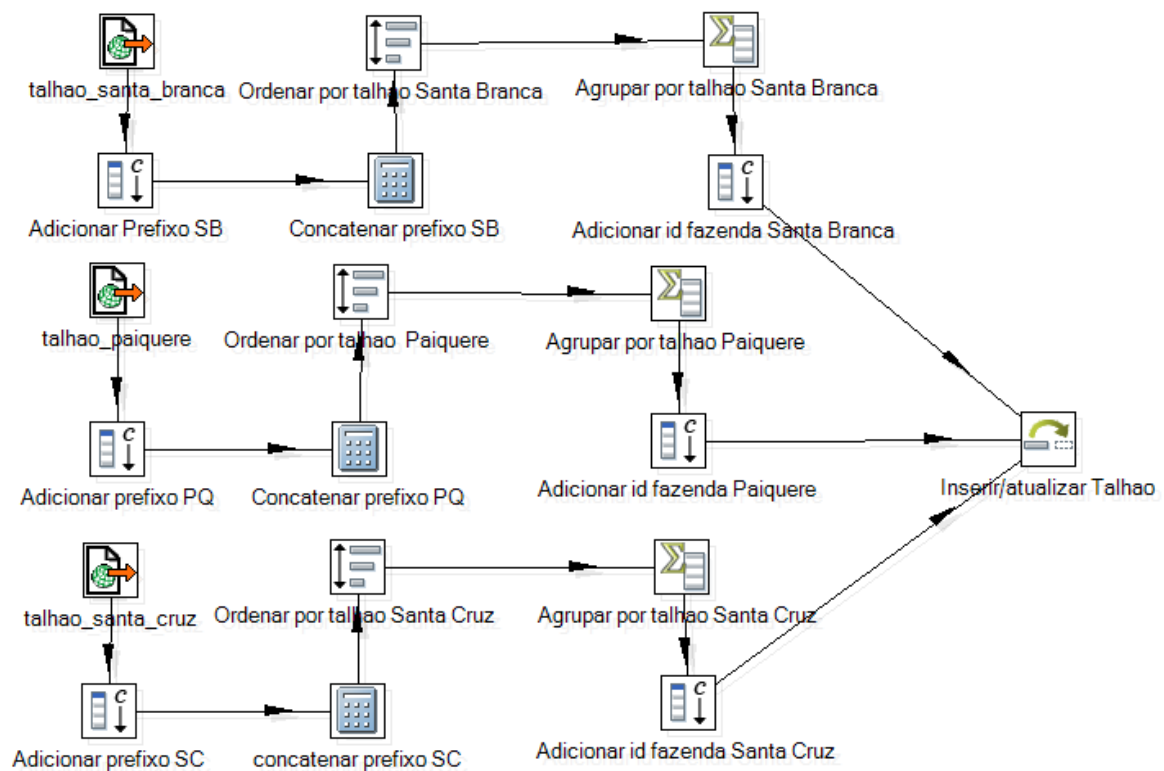


Figura 19: Processo de ETL modelado para inserir hierarquia talhões na dimensão espacial

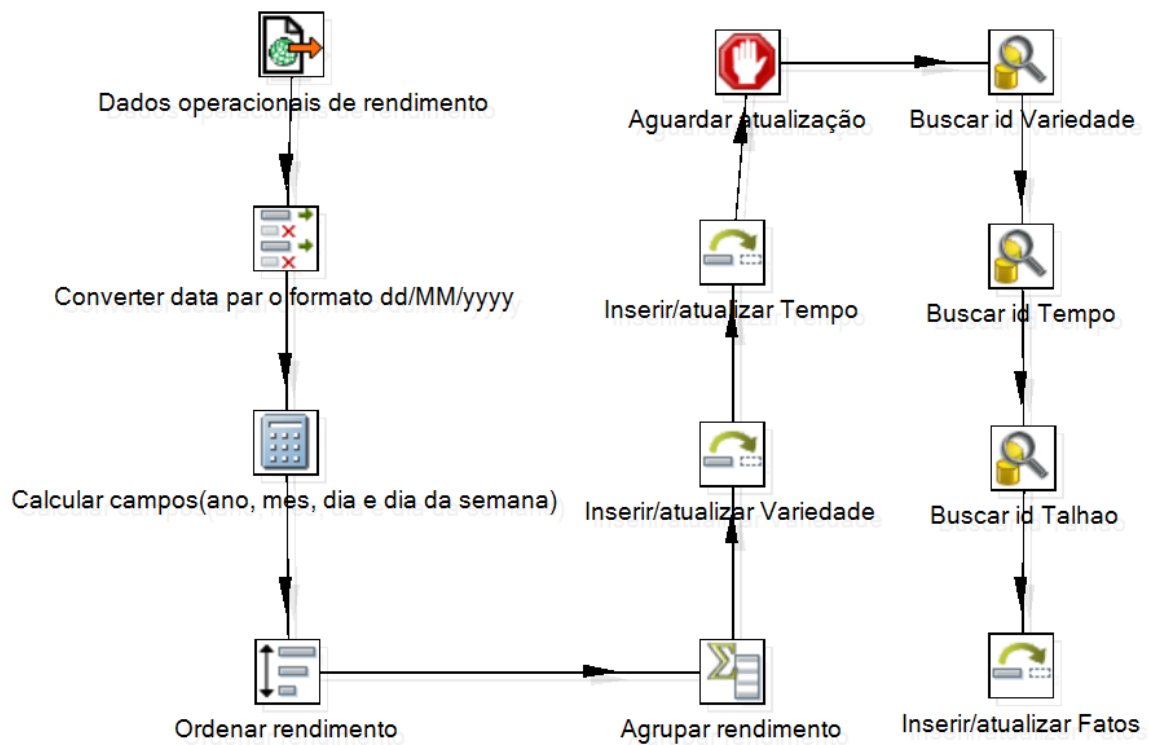


Figura 20: Processo de ETL modelado para inserir fatos de rendimento, buscando as chaves de suas respectivas dimensões

- mostrar uma nuvem de palavras (tipo de gráfico que não existe em dc.js)

Outra contribuição foi a implementação das funcionalidade de Crossfilter para cálculo de média, que não é implementado nativamente.

6.1.3 Interface

A Figura 21, mostra a interface implementada pelo metodo proposto de forma geral. Cada item dos componentes gráficos é um item ativo, que permite o filtro de dados, por meio do evento *click* (ao clicar sobre o item, aplica-se o filtro "where item_clicado"). Assim também é implementado o método de *Drill-down* e *Roll-up* sobre o evento do *scroll* do mouse (*scroll* para baixo *Drill-down* e *scroll* para cima *Roll-up*). Também são calculados para cada dimensão uma escala de cores, baseada nos valores máximos e mínimos. Os componentes podem ser caracterizados como:

- (a): Botões para seleção de cubo e medidas. A interface permite a adição de n cubos para serem analisados. bem como selecionar a medida de estudo da referente análise. Estes dados são recuperados por meio dos métodos descritos na Seção 3.7 (Tecnologias Abertas e Extensíveis Aplicadas ao Suporte a Decisão na Agricultura), os quais consultam o arquivo de metadados elaborado como descrito na Seção ??;
- (b): As nuvens de palavras são construídas baseadas na escala de cores construída para a referida dimensão, sendo as cores de intensidade mais forte, valores numéricos maiores;
- (c): O mapa é uma exclusividade das dimensões que possuem uma propriedade configurada no arquivo referido na Sessão 3.7 como "*Geometry*". Também é utilizado a paleta de cores criada para a dimensão;
- (d): Gráfico de barras sobre a dimensão tempo (timeline) onde podem ser navegado pelos níveis temporais pelas operações *Drill-down* e *Roll-up*;
- (e): O gráfico de pizza, também apresentado em uma paleta de cores, permitido análises de representatividade em relação ao total;
- (f): Gráfico de barra da dimensão espaço, utilizado em conjunto com o mapa, afim de oferecer diferentes perspectivas dos dados;
- (g): E por fim os dados tabulados, informando de forma simplificada os dados numéricos.

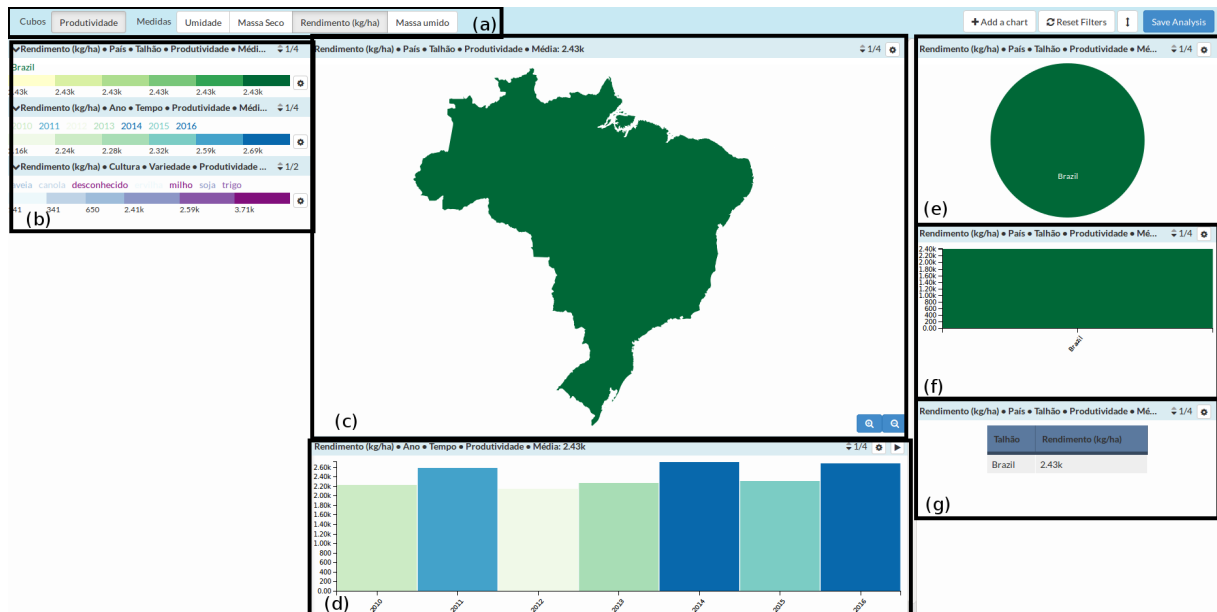


Figura 21: Interface da implementação do método proposto: (a) Seleção de cubos e medidas disponíveis; (b) lista de dimensões dispostas em intensidade de cor; (c) mapa da dimensão espacial; (d) gráfico de barra da dimensão temporal; (e) gráfico de pizza da dimensão espacial; (f) Gráfico de barra da dimensão espacial; e (g) Tabela de medidas

As operações de *Drill-down* e *Roll-up*, são facilitadas através dos eventos de *scroll* do mouse, tornando a navegação intuitiva e fácil. A Figura 22 mostra duas operações de *Drill-down* executadas em sequencia. Após uma operação, todos os itens gráficos(gráfico de coluna, gráfico de pizza, mapa, tabela) são atualizados e redesenhados em tempo de execução. Essa característica é importante, pois a composição de uma análise de um cenário, deve ser rápida e dinâmica, afim de permitir o utilizador navegar fluidamente pelos dados e informações.

Outra operação de navegação que foi facilitada é a operação de *slice* e *dice*, que pode ser realizada de forma dinâmica apenas clicando sobre os itens das dimensões que deseja que seja aplicado um filtro. Essas operações podem ser observadas na Figura 23.

A Figura 23 pode ser utilizada em um estudo de caso, como prova de conceito, onde obtém-se exemplos de análises: não foram todos os talhões que produziram milho no ano de 2015, também nota-se que os talhões mantiveram uma média de produtividade, apenas destacando dois casos, um positivamente e outro negativamente. Estas cenários de análise devem ser utilizados em conjunto com demais propriedades e atributos que podem ter influenciado na produtividade, e realizar correlações entre os manejos de cultura em busca do melhor rendimento, bem como a redução de custos.

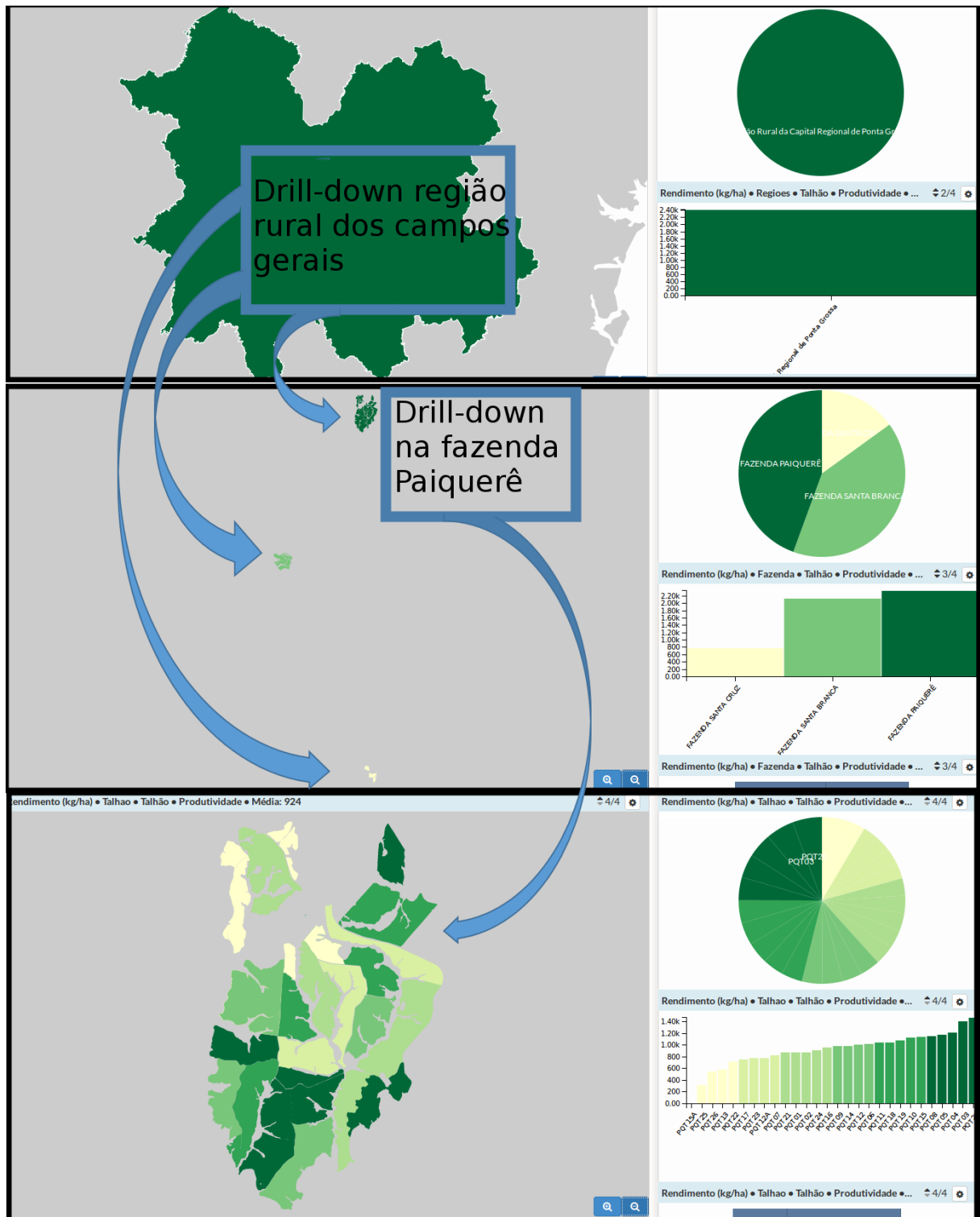


Figura 22: Operações de *Drill-down* sucessivas e o comportamento dos gráficos e tabelas

6.2 Considerações

Neste capítulo foi apresentado os resultados desta pesquisa, sendo: Uma arquitetura para a integração de ambientes *Data Warehouse*, SIG e agricultura, sendo que abrange, desde o baixo nível de banco de dados, até o alto nível da interface com o usuário. Também foi abordado

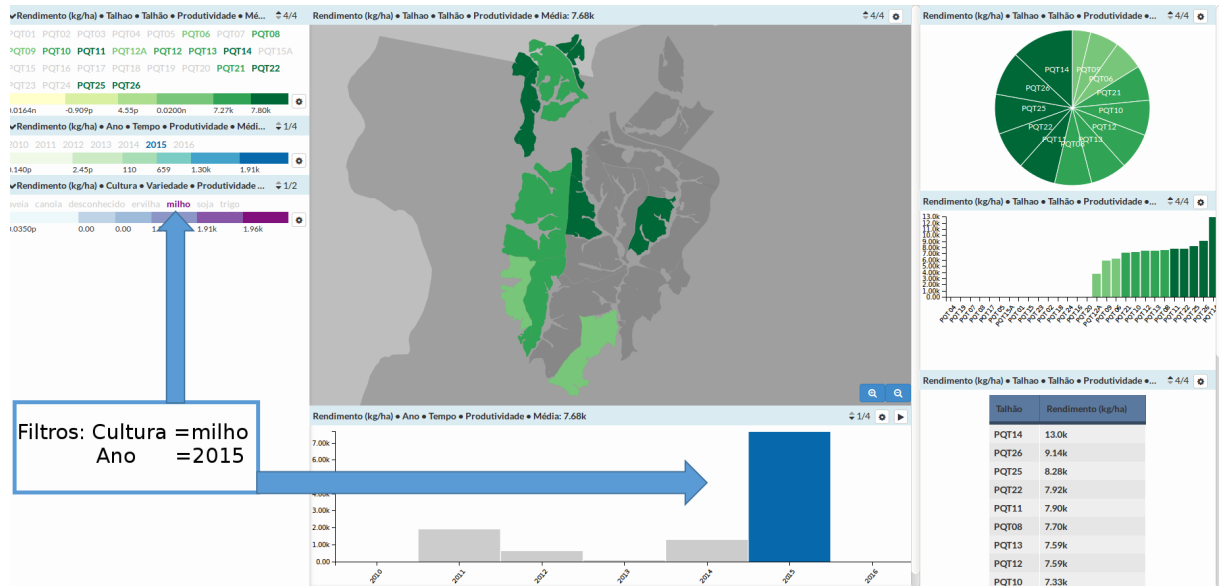


Figura 23: Operações de filtro para selecionar a produção de milho no ano de 2015, para os talhões da fazenda Paiquerê

o processo ELT, sendo descrito todas as etapas, transformações e conversões necessárias.

No próximo capítulo será apresentado as considerações finais desta pesquisa, destacando os objetivos alcançados, bem como algumas dificuldades e trabalhos futuros.

Considerações Finais

O processo de tomada de decisão agrícola não é simples, pois a eficácia de uma decisão é dependente de muitos fatores, sendo inúmeros deles incontroláveis, com por exemplo, temperatura, chuvas, umidade, pragas, ervas daninhas, entre outros. Entretanto, ferramentas que promovem o apoio, e gerência de informações tem um papel importante, afim auxiliar o aprendizado gerencial de uma área produtiva. É neste ponto que se encaixam as ferramentas de apoio a decisão como *Data Warehouse* e OLAP. Porém dados agrícolas, possuem variabilidade espacial, sendo este atributo essencial para contribuir com as conclusões analíticas que se pode obter sobre os dados agrícolas.

A integração entre os conceitos e tecnologias não é trivial, necessitando um estudo aprofundado e a elaboração de métodos que permitam unir as vantagens de cada conceito (*Data Warehouse*, SIG e Agricultura). Neste trabalho, foram analisados frameworks e bibliotecas existentes como por exemplo geonondrian, geonode, DC.js, entre outros que proporcionam integrar os conceitos citados, de forma simples eficiente e rápida. E também foi modelado com base na topologia estrela um sistema agrícola a partir de dados exportados de sensores de colhedoras. Esses sensores coletam informações como GPS, umidade, quantidade, velocidade, altitude, largura.

Outro ponto abordado nesta dissertação são os protocolos de comunicação entre as diversas camadas do processo de integração. Para isso foi utilizados os protocolos XMLA sobre

http, Json sobre *socket* e Json sobre HTTP.

Como principal contribuição, foi proposta uma arquitetura de software capaz de suportar a integração proposta, e prover uma aplicação para análise de dados multidimensionais com atributos geográficos.

A camada de apresentação constitui um item importante, pois é o meio de acesso do usuário final, com toda a arquitetura e conceitos desenvolvidos nesta dissertação. Para tanto, ela foi desenvolvida afim de facilitar e agilizar o processo analítico, por meio de cliques e rolagem do mouse. E por fim a utilização de tecnologias abertas e extensíveis, onde foram utilizados desde as ferramentas para o processo ETL, até os componentes da arquitetura de código aberto. Proporcionando assim a redução de custos, fácil aplicação, extensibilidade e o reuso de soluções existentes. Por consequência, a solução desenvolvida, também compartilha das características de extensibilidade, e está disponibilizada de forma aberta por meio do repositório público github.

As principais dificuldades encontradas foram em relação a obtenção dos dados. A agricultura de precisão no Brasil tem muito para se desenvolver para ser uma área consolidada, uma vez que se faz necessário a modernização dos equipamentos, logo é necessário investimentos, que muitas vezes são pesados para o orçamento do produtor brasileiro. Também foi encontrado dificuldades para a limpeza dos dados. Para esta dissertação, a limpeza constituiu de eliminar dados nulos de sensores. Porém, ainda nota-se durante o processo de análise dados não condizentes com a realidade.

Como trabalhos futuros, sugere-se o refinamento do processo ETL, por meio de equacionamento de uma função de rendimento, permitindo a identificação de anomalias em dados oriundos de sensores, bem como a possível correção de valores e preenchimento de buracos (*gaps*) nos dados. A validação desta arquitetura, por meio de dados reais, e aplicação para apoio a decisões em cenários reais.

REFERÊNCIAS

- ARAÚJO, A. L. de. *A tecnologia da geoinformação como ferramenta para a modernização da gestão municipal em cidades de pequeno porte*. Dissertação (Mestrado) — Universidade Federal de Pernambuco, 2003.
- ARONOFF, S. Geographic information system: a management perspective. *WBL Publications*, 1989.
- BADARD, T. Geokettle: A powerful open source spatial etl tool. *OSS4G 2009 conference*, 2009.
- BARROS, D. M. V. *AGIS: Um Serviço para Processamento Geográfico e Analítico Multinível*. Dissertação (Mestrado) — Universidade Federal de Pernambuco, 2009.
- BÉDARD, Y.; MERRETT, T.; HAN, J. Fundamentals of spatial data warehousing for geographic knowledge discovery. *Geographic Data Mining and Knowledge Discovery*, 2001.
- BIMONTE, S.; WEHRLE, P. Gewolap: A web based spatial olap proposal. In: *Second International Workshop on Semantic based Geographical Information Systems*. [S.l.: s.n.], 2006.
- BOSTOCK, M. *Data-Driven Documents (D3.js)*. [S.l.]: GitHub, 2016. <https://github.com/d3/d3/wiki>. [Online; Acesso em: 22 jul. 2016].
- BURROUGH, P. A. Principles of geographical information systems for land resources assessment. *Clarendon Press*, 1986.
- BURROUGH, P. A.; MCDONNELL, R. A. Principles of geographical information systems: Spatial information systems. *Oxford University Press*, 1998.
- CÂMARA, G. *et al.* Anatomia de sistemas de informações geográficas. *UNICAMP: Instituto de Computação*, 1996.
- CROCKFORD, D. *The application/json Media Type for JavaScript Object Notation (JSON)*. IETF, jul. 2006. RFC 4627 (Informational). (Request for Comments, 4627). Disponível em: <<http://www.ietf.org/rfc/rfc4627.txt>>.

- DC.js. *Dimensional Charting Javascript Library (DC.js)*. [S.l.]: GitHub, 2016. <https://github.com/d3/d3/wiki>. [Online; Acesso em: 22 jul. 2016].
- Deere & Company. *APEX software para gerenciamento de fazendas*. 2011. https://www.deere.com.br/common/docs/stellar/es_R3/Agriculture/Ag_Management_Solutions/Software/pdf/ApexUserGuide_PT.pdf. [Online; Acesso em: 10 nov. 2016].
- DEMERS, M. N. *Fundamentals of Geographic Information Systems*. 4. ed. [S.l.]: John Wiley & Sons, Inc., 2008.
- ELMASRI, R.; NAVATHE, S. B. *Fundamentals of Database Systems*. [S.l.]: Pearson Addison Wesley, 2010.
- Environmental Systems Research Institute, Inc. *ESRI Shapefile Technical Description*. 1998. <https://www.esri.com/library/whitepapers/pdfs/shapefile.pdf>. [Online; Acesso em: 10 nov. 2016].
- FIELDING, R. *et al.* *Hypertext Transfer Protocol – HTTP/1.1*. United States: RFC Editor, 1999.
- FOUNTAS, S. *et al.* Farm management information systems: Current situation and future perspectives. *Computers and Electronics in Agriculture*, v. 115, p. 40–50, 2015.
- FRAZIER, B. E.; WALTERS, C. S.; PERRY, E. M. *The State of Site-Specific Management for Agriculture*. [S.l.]: American Society of Agronomy, Crop Science Society of America, and Soil Science Society of America, 1997.
- FU, L.; HAMMER, J. *CUBIST: A New Approach to Speeding Up OLAP Queries in Data Cubes*. [S.l.], 2001.
- GEBBERS, R.; ADAMCHUK, V. Precision agriculture and food security. *Science Magazine*, 2010.
- GLORIO, O. *et al.* A personalization process for spatial data warehouse development. *Decision Support Systems*, v. 52, p. 884–898, 2012.
- GOMES, E. A. *Arquitetura, Metamodelos e Algoritmos para um Engenho SOLAP*. Dissertação (Mestrado) — Universidade Federal de Pernambuco, 2011.
- GOOGLE. *Google Maps APIs*. 2016. <https://developers.google.com/maps/documentation/>. [Online; Acesso em: 18 mar. 2016].
- HEEGE, H. J.; REUSCH, S.; THIESSEN, E. Prospects and results for optical systems for site-specific on-the-go control of nitrogen-top-dressing in germany. *Precision Agriculture*, v. 9, p. 115–131, 2008.
- HYDE, J. *Mondrian Documentation*. 2016. <http://mondrian.pentaho.com/documentation/olap.php>. [Online; Acesso em: 18 mar. 2016].
- HYDE, J.; KLAWANS, B. *olap4j Specification*. 2011. http://www.olap4j.org/olap4j_fs.html. [Online; Acesso em: 18 mar. 2016].
- INMON, H. W. *What is a data warehouse*. [S.l.]: Prism Tech Topic, 1995.

- Instituto Brasileiro de Geografia e Estatística. *Regiões Rurais*. 2015. http://www.ibge.gov.br/home/geociencias/geografia/default_regioes_rurais.shtm. [Online; Acesso em: 10 nov. 2016].
- KIMBALL, R.; ROSS, M. *The Data Warehouse Toolkit*. 3. ed. [S.l.]: John Wiley & Sons, Inc., 2013.
- KYUNG, M.-J.; YOM, J.-H.; KIM, S.-Y. Spatial data warehouse design and spatial olap implementation for decision making of geospatial data update. *Journal of Civil Engineering*, v. 16, p. 1023–1031, 2012.
- MCBRATNEY, A.; WHELAN, B.; ANCEV, T. Future directions of precision agriculture. *Precision Agriculture*, v. 6, p. 7–23, 2005.
- MCGUIRE, M. *et al.* A user-centered design for a spatial data warehouse for data exploration in environmental research. *Ecological Informatics*, v. 3, n. 4?5, p. 273 – 285, 2008. ISSN 1574-9541. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1574954108000496>>.
- MELTON, J.; SIMON, A. R. Understanding the new sql: A complete guide. *Morgan Kaufmann*, 1993.
- MICROSOFT. *Analysis Services*. 2016. <http://technet.microsoft.com/ptbr/library/bb522607.aspx>. [Online; Acesso em: 18 mar. 2016].
- Microsoft and Hyperion Solutions. *XML for Analysis Specification*. 2002. <http://www.xmlforanalysis.com/xmla1.1.doc>. [Online; Acesso em: 10 nov. 2016].
- MURAKAMI, E. *et al.* An infrastructure for the development of distributed service-oriented information systems for precision agriculture. *Computers and Electronics in Agriculture*, v. 58, p. 37–48, 2007.
- NIKKILA, R.; SEILONEN, I.; KOSKINEN, K. Software architecture for farm management information systems in precision agriculture. *Computers and Electronics in Agriculture*, v. 70, p. 328–336, 2010.
- OGC. *OpenGIS Geography Markup Language (GML) Encoding Standard*. [S.l.]: Open Geospatial Consortium, 2007.
- Open Source Geospatial Foundation. *GeoNodes Documentation*. 2016. <http://docs.geonode.org/en/master/>. [Online; Acesso em: 10 nov. 2016].
- ORACLE. *Oracle Fusion Middleware MapViewer*. 2016. <http://www.oracle.com/technetwork/pt/middleware/mapviewer/overview/index.html>. [Online; Acesso em: 18 mar. 2016].
- ORACLE. *Spatial and Graph Developer's Guide*. 2016. <http://docs.oracle.com/database/121/SPATL/toc.htm>. [Online; Acesso em: 18 mar. 2016].
- OSGEO. *PostGIS*. 2016. <http://postgis.net/>. [Online; Acesso em: 19 mar. 2016].
- OSGEO Project. *PostGIS 2.0 Manual*. 2016. <http://postgis.net/docs/manual-2.0/>. [Online; Acesso em: 10 nov. 2016].

- PAIM, F. R. S. *Uma Metodologia para Análise de requisitos em Sistemas Data Warehouse*. Dissertação (Mestrado) — Cin \UFPE, 2003.
- PAULINO, L. A.; CARNEIRO, A. F. T. Base de dados gráficos para sistemas de informações geográficas (sig?s). In: MULTIFINALITÁRIO, C. B. D. C. T. (Ed.). *Anais do III Congresso Brasileiro de Cadastro Técnico Multifinalitário*. [S.l.: s.n.], 1998.
- RAO, S. G. Providing better support for a class of decision support queries. In: ACM-SIGMOD (Ed.). *Int'l Conf. on Management of Data*. [S.l.: s.n.], 1996.
- ROBERT, T. *Mandoline*. [S.l.]: GitHub, 2014a. <https://github.com/ThomasRobertFr/mandoline>. [Online; Acesso em: 22 jul. 2016].
- ROBERT, T. *Crossfilter server*. [S.l.]: GitHub, 2014c. <https://github.com/ThomasRobertFr/crossfilter-server>. [Online; Acesso em: 22 jul. 2016].
- ROBERT, T. *Query API*. [S.l.]: GitHub, 2014d. <https://github.com/ThomasRobertFr/solap4py-js>. [Online; Acesso em: 22 jul. 2016].
- SILVA, J. da *et al.* Modelling and querying geographical data warehouses. *Information Systems*, v. 35, p. 592–614, 2010.
- SMITH, M. J. de; GOODCHILD, M. F.; LONGLEY, P. A. *Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools*. [S.l.]: The Winchelsea Press, Winchelsea, UK, 2015.
- SORENSEN, C. *et al.* Functional requirements for a future farm management information system. *Computers and Electronics in Agriculture*, v. 76, p. 266–276, 2011.
- SPATIALYTICS. *GeoMondrian Documentation*. 2016. http://wiki.spatialytics.org/doku.php?id=en:spatialytics_olap. [Online; Acesso em: 18 mar. 2016].
- STAFFORD, J. V. Implementing precision agriculture in the 21st century. *Journal of Agricultural Engineering Research*, v. 76, p. 267–275, 2000.
- THOMSEN, E. *OLAP Solutions: Building Multidimensional Information Systems*. 2. ed. [S.l.]: John Wiley & Sons, Inc., 2002.
- TONBELLER, A. *Web Component Framework*. 2016. <http://jpivot.sourceforge.net/wcf/index.html>. [Online; Acesso em: 10 nov. 2016].
- VERNIER, F. *et al.* Eis pesticides: An environmental information system to characterize agricultural activities and calculate agro-environmental indicators at embedded watershed scales. *Agricultural Systems*, v. 122, p. 11–21, 2013.

APÊNDICE A – INSTALAÇÃO DE BIBLIOTECAS E MÓDULOS

A.1 Exportar arquivos do software APEX

Para exportar os dados de rendimento e os limites de fazenda e talhão por meio do software de gerenciamento APEX (Deere & Company, 2011), foi utilizado o seguinte procedimento:

1. Selecione "Arquivo > Exportar Mapa > Exportar Lote".
2. Na caixa "Opções de Exportação" localizada no lado esquerdo superior da tela selecione o link "Exportar Todos".
3. Na opção "Exportar Todos" o Apex possui um filtro de exportação acima da árvore de mapas que permite filtrar os dados que deseja exportar por cliente, operação ou lista de limite.
 - (a) Ao selecionar o filtro "Por Operações" uma segunda caixa de filtro aparecerá à direita, permitindo que você filtre por Aplicação, Semeadura, Cultivo ou Outros.
4. Na janela "Exportar Arquivos" à direita é possível simplificar sua seleção selecionando "Exportar Todos" ou "Exportar Selecionado".
 - (a) Selecione "Exportar Todos" se deseja exportar todos os mapas disponíveis
 - (b) Selecione "Exportar Selecionado" para selecionar os dados específicos que deseja exportar.
5. Se selecionar "Exportar Todos" prossiga no passo 7. Se selecionar "Exportar Selecionado" prossiga no próximo passo.
6. O Apex lhe dará duas Opções de Exportação.
 - (a) Compactado – Quando marcado, o Apex compactará seus dados exportados em um formato de arquivo .zip.
 - (b) Dados de Calibração – Quando marcado, o Apex exportará os dados calibrados. Se essa caixa estiver desmarcada, o Apex exportará os dados brutos.
7. Na caixa "Exportar Tipos de Arquivo" selecione o formato de arquivo "CSV ou Shape".
 - (a) Arquivo CSV – formato de arquivo texto que contém valores delimitados por vírgula.
 - (b) Shapefile – formato padrão do arquivo ESRI com ponto, polígono e linhas para exibir um mapa do Sistema de Informações Geográficas

8. Na caixa "Salvar na Pasta" clique no botão "Reticências" para selecionar o local para os dados serem salvos.

9. Selecione o local para salvar seus dados e selecione "OK".

10. Selecione "Exportar" para exportar seus dados.

A.2 Instalação do SGBD PostgreSQL e a extensão PostGIS

```
# Instalando o PostgreSQL, PostGIS e o PgAdmin:
$ sudo apt-get install postgresql-9.1 postgresql-9.1-postgis-2.0 pgadmin3

#Troque o seu usuário para o usuário postgres com os comandos
$ sudo su
$ su postgres

#E crie o banco de dados postgis:
$ createdb postgis

#Em seguida, crie um usuário para acessar o SGBD com o comando
#createuser nome_do_usuario:
$ createuser agribiuser -P
Enter password for new role:
Enter it again:
Shall the new role be a superuser? (y/n) y

$ su agribiuser

#Acesse o banco e digite a instrução CREATE EXTENSION postgis;
$ psql postgis
psql (9.1.10)
Type "help" for help.

postgis=# CREATE EXTENSION postgis;

#Para criar bancos de dados espaciais a partir do banco postgis criado
#anteriormente, utilize o comando createdb com a flag -T (template)
```

```
$ createdb agribi -T postgis
```

A.3 Instalação e configuração GeoMondrian

A.3.1 Pré requisitos

1. Java Runtime Environment (JRE)
2. Servlet container como o Apache Tomcat
3. SGBD PostgreSQL com a extensão espacial (PostGIS) ativada

A.3.2 Instalação

1. Download do projeto geomondrian (SPATIALYTICS, 2016)
2. Deploy do arquivo WAR no servelet container

A.3.3 Configuração

1. Configurar o arquivo datasources.xml no diretório <dir de instalacao>/WEB-INF/

```
<DataSourceInfo>
    Provider=mondrian;
    Jdbc=jdbc:postgresql_postGIS:
        //localhost:5432/agribi;
    JdbcDrivers=org.postgis.DriverWrapper;
    JdbcUser=agribiuser;
    JdbcPassword=123456;
    PoolNeeded=false;
    Catalog=/WEB-INF/queries/agribi.xml
</DataSourceInfo>
```

2. Criar o arquivo xml de configuração agribi.xml no diretório <dir de instalacao>/WEB-INF/queries/

Código A.1: XML de metadados para servidor de consulta OLAP Espacial

```
1 <Schema name="cubo">
2 <Cube name="Produtividade" cache="true" enabled="true">
3 <Table name="fato" schema="public">
```

```

4      </Table>
5      <Dimension foreignKey="id_talhao" name="Talhao">
6      <Hierarchy hasAll="true" primaryKey="id_talhao"
          primaryKeyTable="dim_talhao">
7      <Join leftKey="id_fazenda" rightAlias="dim_fazenda" rightKey="
          id">
8      <Table name="dim_talhao">
9      </Table>
10     <Join leftKey="regiao_id" rightKey="regiao_id">
11     <Table name="dim_fazenda">
12     </Table>
13     <Table name="dim_regioes">
14     </Table>
15     </Join>
16     </Join>
17     <Level name="Regioes" table="dim_regioes" column="nome" type="
        String" uniqueMembers="true" levelType="Regular"
        hideMemberIf="Never">
18     <Property name="Geom" column="geom" type="Geometry">
19     </Property>
20     </Level>
21     <Level name="Fazenda" table="dim_fazenda" column="nome" type="
        String" uniqueMembers="true" levelType="Regular"
        hideMemberIf="Never">
22     <Property name="Geom" column="geom" type="Geometry">
23     </Property>
24     </Level>
25     <Level name="Talhao" table="dim_talhao" column="talhao" type="
        String" uniqueMembers="false" levelType="Regular"
        hideMemberIf="Never">
26     <Property name="Geom" column="geom" type="Geometry">
27     </Property>
28     </Level>
29     </Hierarchy>
30     </Dimension>
31     <Dimension name="Tempo" type="TimeDimension" foreignKey="
        id_tempo">
32     <Hierarchy hasAll="false" primaryKey="id_tempo">
33     <Table name="dim_tempo"/>
34     <Level name="Ano" column="ano" type="Numeric" uniqueMembers="
        true"
35     levelType="TimeYears"/>
36     <Level name="Mes" column="mes" uniqueMembers="false" type="

```

```

        Numeric"
37         levelType="TimeMonths"/>
38         <Level name="Dia" column="dia" uniqueMembers="false" type="
        Numeric"
39         levelType="TimeDays"/>
40     </Hierarchy>
41     <Hierarchy hasAll="true" name="Semanalmente" primaryKey="
        id_tempo">
42         <Table name="dim_tempo"/>
43         <Level name="Ano" column="ano" type="Numeric" uniqueMembers="
        true"
44         levelType="TimeYears"/>
45         <Level name="Semana" column="dia_da_semana" type="Numeric"
        uniqueMembers="false"
46         levelType="TimeWeeks"/>
47     </Hierarchy>
48 </Dimension>
49 <Measure name="Rendimento_Seco" column="rendimento_seco"
        datatype="Numeric" aggregator="sum" visible="true">
50 </Measure>
51 <Measure name="Rendimento_umido" column="rendimento_umido"
        datatype="Numeric" aggregator="sum" visible="true">
52 </Measure>
53 <Measure name="Umidade" column="umidade" datatype="Numeric"
        aggregator="sum" visible="true">
54 </Measure>
55 </Cube>
56 </Schema>

```

A.4 Instalação API SOLAP

```

# Instalacao de ferramentas de compilacao
sudo apt-get install ant
sudo apt-get install openjdk-7-jdk

# Recuperar o codigo fonte do repositorio
git clone https://github.com/petroskilp/mandoline.git

#Configurar o banco de dados
dbhost=localhost

```

```
dbport=8080
# Driver para acessar a API olap4j
driverName=org.olap4j.driver.xmla.XmlaOlap4jDriver

# Tornar a variavel ambiente JAVA_HOME com Java 7 JDK:
export JAVA_HOME=/usr/lib/jvm/java-7-openjdk-amd64/

# Construir e rodar a aplicacao
ant run
```

A.5 Instalação Geonode

1. Configurar o repositório PPA do Geonode

```
sudo add-apt-repository ppa:geonode/stable
```

2. Instalar o pacote geonode e suas dependências

```
sudo apt-get update; sudo apt-get install geonode
```

3. Configurar o endereço IP (o comando abaixo busca automaticamente)

```
IP_ADDRESS=$(ip route get 8.8.8.8 | awk '{print $NF; exit}') sudo geonode
```

Observação: O endereço ip deve estar habilitado para acesso de outra máquina.

4. Criar um super usuário geonode, o qual será usado na aplicação web.

```
geonode createsuperuser
```

5. Acessar a aplicação por meio do web browser, utilizando o endereço ip configurado no passo 3.

A.6 Instalação da extensão analytics e configuração

```
#recuperar o codigo fonte no repositorio
git clone https://github.com/petroskilp/analytics.git
cd analytics
```

```
# sincronizar com a base do geonode e adicionar a extensao
python manage.py syncdb

# No diretorio do geonode executar o seguinte comando
paver start_geoserver

# para iniciar o Geoserver (funcionalidades de camadas)

# no diretorio do analytics
python manage.py runserver
```