RICARDO AUGUSTO GOMES DA COSTA

UTILIZAÇÃO DE DATA WAREHOUSES PARA GERENCIAR DADOS DE REDES DE SENSORES SEM FIO QUE MONITORAM POLINIZADORES

São Paulo
2011

RICARDO AUGUSTO GOMES DA COSTA

UTILIZAÇÃO DE DATA WAREHOUSES PARA GERENCIAR DADOS DE REDES DE SENSORES SEM FIO QUE MONITORAM POLINIZADORES

Dissertação apresentada à Escola Politécnica da Universidade de São Paulo para obtenção do Título de Mestre em Engenharia Elétrica.

Área de Concentração:
Sistemas Digitais

Orientador: Professor Livre Docente
Carlos Eduardo Cugnasca

São Paulo
2011

Este exemplar foi revisado e alterado em relação à versão original, sob responsabilidade única do autor e com a anuência de seu orientador.

São Paulo, 26 de outubro de 2011.

Assinatura do autor _____

Assinatura do orientador _____

FICHA CATALOGRÁFICA

Costa, Ricardo Augusto Gomes da
Utilização de data warehouses para gerenciar dados de redes de sensores sem fio que monitoram polinizadores / R.A.G. da Costa. -- ed. rev. -- São Paulo, 2011.
78 p.

Dissertação (Mestrado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Computação e Sistemas Digitais.

1.Sistemas de comunicação sem fio 2.Modelagem de dados 3.Sistemas de informação 4.Abelhas I.Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Computação e Sistemas Digitais II.t.

DEDICATÓRIA

Aos meus pais, Ayrton e Zélia que sempre incentivaram e apoiaram minha formação profissional e, principalmente, minhas decisões.

AGRADECIMENTOS

Ao meu orientador, Carlos Cugnasca, pelo apoio, conselhos e atenção dispensada durante essa jornada.

Aos colegas do MINTER UEA-USP, pelo companheirismo, amizade, discussões e apoio mútuo.

Aos colegas do LAA, pelo apoio e parceria.

Ao colega Ayrton Vollet Neto, mestrando em Biologia da USP de Ribeirão Preto, que ajudou com informações técnicas sobre abelhas.

À Suframa pelo patrocínio financeiro ao projeto MINTER UEA-USP.

À Capes pela criação e regulamentação do mestrado interinstitucional.

À UEA por gerar as condições e propiciar o oferecimento do programa de Pós da Escola Politécnica da Universidade de São Paulo em Manaus.

À Fundação de Amparo à Pesquisa do Estado do Amazonas (FAPEAM), que por meio da bolsa RH-INTERINSTITUCIONAL permitiu a realização de estágio na Escola Politécnica da USP em São Paulo.

À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), pelo apoio ao projeto TIDIA/VINCES.

E por fim agradeço a todos que me ajudaram de forma direta ou indireta para a realização deste trabalho.

RESUMO

Este trabalho tem como objetivo a aplicação do conceito de *data warehouse* para a agregação, gerenciamento e apresentação de dados coletados por meio de Redes de Sensores Sem Fio que monitoram polinizadores. Os experimentos científicos que utilizam tais redes para monitorar *habitat* geram um volume de dados que precisa ser tratado e analisado, para que possa auxiliar os pesquisadores e demais interessados nas informações. Tais dados, gerenciados e correlacionados com informações de outras fontes, podem contribuir para a tomada de decisões e ainda realimentar outros experimentos. Para a avaliação da proposta, desenvolveu-se um modelo para extração, transformação e normalização dos dados coletados por redes de sensores sem fio, contemplando ainda a carga em *data warehouse*. Considerou-se no modelo, dados tabulados das redes de sensores sem fio, utilizados em experimentos com abelhas e ainda dados de outras fontes sobre o cultivo de abelhas, importantes para obtenção de visões do *data warehouse* mais apuradas. O uso de *data warehouse* aplicado a esse contexto mostrou-se um alternativa viável e útil, pois facilitou a obtenção de dados consolidados sobre o experimento, importante para a tomada de decisão pelos pesquisadores e ainda, diminui o tempo gasto pelos interessados em extrair essas informações, em comparação à tradicional análise em planilhas eletrônicas.

Palavras-chave: monitoração de *habitat*, data warehouse, redes de sensores sem fio.

ABSTRACT

This work aims at applying the concept of data warehouse for data aggregation, management and presentation of data collected by Wireless Sensor Networks that monitor pollinators. Scientific experiments using such networks to monitor *habitat* generate a volume of data that must be addressed and analyzed, so that they can help researchers and others interested in the information. This data, managed and correlated with information from other sources may contribute to the making and still replenish other experiments. For the evaluation of the proposal, it was developed a model for the extraction, processing and standardization of data collected by wireless sensor networks, covering also the load on the data warehouse. It was considered in the model tabulated data networks of wireless sensors, used in experiments with bees and even data from other sources on the cultivation of bees, important to obtain views of the data warehouse more accurate. The use of data warehouse implemented in this context proved to be a viable and useful, as it facilitated the obtaining of information for decision making by researches and stakeholders and reduces time consumed by stakeholders to extract such information.

Keywords: *Habitat* monitor, data warehouse, wireless sensor network.

LISTA DE FIGURAS

Figura 1. Arquitetura de um nó sensor. Adaptada de (AKYILDIZ, SU, *et al.*, 2002). . 20
Figura 2. Exemplo de um RSSF. 22
Figura 3. Placa MPR2400. Extraído de (MEMSIC, 2010). 27
Figura 4. MIB 520 – *Gateway* USB. Extraído de (MEMSIC, 2010). 27
Figura 5. *Data Warehouse* orientado a assuntos. 29
Figura 6. Dados sobre temperatura sendo convertidos antes de serem colocados no DW. 30
Figura 7. Exemplo de modelo dimensional com três dimensões e uma tabela fato. . 35
Figura 8. Esquema estrela (*star schema*). 36
Figura 9. Esquema floco de neve (*snow flake schema*). 37
Figura 10. Processo ETL em um *Data Warehouse.* 38
Figura 11. Exemplo de matriz de barramento. 39
Figura 12. Arquitetura proposta. 45
Figura 13. Dados INPE - Temperatura média 2010 - São Paulo. 48
Figura 14. Fotos de nós sensores coletando dados. Extraído de (NETO, 2010). 53
Figura 15: Extração, Transformação e Carga dos dados coletados. 54
Figura 16. Modelo dimensional proposto. 56
Figura 17. Matriz de barramento do modelo proposto. 56
Figura 18. Tela do protótipo criado utilizando o Mondrian/Pentaho. 57
Figura 19. Planilha eletrônica gerada pelo Mondrian/Pentaho. 58
Figura 20. Gráfico gerado pelo Mondrian/Pentaho. 58
Figura 21: Tela do protótipo com visão customizada. 58
Figura 22. Tela de entrada Mondrian/Pentaho. 71
Figura 23. Tela inicial do Mondrian/Pentaho. 71
Figura 24. Tela inicial do Painel de Administração do Mondrian/Pentaho. 72
Figura 25. Configuração da conexão com banco de dados. 73
Figura 26. Tela de análise do Mondrian/Pentaho. 76
Figura 27. Barra de ferramentas do Mondrian/Pentaho 76
Figura 28. Trecho de código do ETL 77
Figura 29. Trecho de código usado para Transformação dos dados. 78

LISTA DE TABELAS

Tabela 1. Exemplo de não volatilidade de dados em DW. 31
Tabela 2. Dados monitorados em experimentos com abelhas. 47
Tabela 3. Exemplo de dados sobre temperatura. Extraído de (NETO, 2010). 52
Tabela 4. Exemplo de dados sobre umidade. Extraído de (NETO, 2010). 52

LISTA DE ABREVIATURAS E SIGLAS

BI	Inteligência Gerencial (*Business Intelligence*)
DSS	Sistema de suporte a decisão (*Decision Support System*)
DW	Depósito de dados (*Data Warehouse*)
ETL	Extração, Transformação e Carga (*Extract, Transform and Load*)
OLAP	Processamento Analítico Online (*Online Analytical Processing*)
RSSF	Redes de Sensores sem Fio

SUMÁRIO

1 INTRODUÇÃO 13
1.1 CONSIDERAÇÕES INICIAIS E MOTIVAÇÃO 13
1.2 OBJETIVO 17
1.3 CONTEÚDO E ORGANIZAÇÃO 18
2 REDES DE SENSORES SEM FIO 19
2.1 INTRODUÇÃO 19
2.2 NÓ SENSOR 20
2.3 GATEWAY 22
2.4 SISTEMA OPERACIONAL TINYOS 23
2.5 PROTOCOLOS DE COMUNICAÇÃO 23
2.5.1 ZigBee 23
2.6 PROTOCOLO DE ROTEAMENTO 24
2.7 CARACTERÍSTICAS 25
2.8 UM EXEMPLO DE REDES DE SENSORES SEM FIO 26
2.9 CONSIDERAÇÕES SOBRE O CAPÍTULO 28
3 DATA WAREHOUSE 29
3.1 INTRODUÇÃO 29
3.2 CARACTERÍSTICAS 32
3.3 OLAP 32
3.4 DSS 33
3.5 DATA MINING 33
3.6 DATA MART 33
3.7 MODELAGEM DIMENSIONAL 34
3.8 ESQUEMAS DIMENSIONAIS 36
3.9 ETL 38
3.10 MATRIZ DE BARRAMENTO 39
3.11 CONSIDERAÇÕES SOBRE O CAPÍTULO 40
4 TRABALHOS RELACIONADOS 41
4.1 MONITORAÇÕES DE ANIMAIS E HABITAT COM RSSF 41
4.2 DATA WAREHOUSES APLICADOS A MONITORAÇÃO DE ANIMAIS 43

4.3 GERENCIAMENTO DE DADOS COLETADOS POR RSSF 43
4.4 CONSIDERAÇÕES SOBRE O CAPÍTULO 44
5 ARQUITETURA PROPOSTA 45
5.1 VISAO GERAL.....45
5.2 POR QUE RSSFs?.....46
5.3 OUTROS SISTEMAS (OUTRAS FONTES).....48
5.4 INTEGRAÇÃO DOS DADOS.....49
5.5 CONSIDERAÇÕES SOBRE O CAPÍTULO 49
6 CASO DE USO..... 50
6.1 PROJETO VINCES 50
6.2 TERMORREGULAÇÃO.....51
6.3 PROTÓTIPO FUNCIONAL 51
6.3.1 Modelagem dos dados..... 51
6.3.2 Extração, Transformação e Carga dos dados..... 53
6.3.3 Dados de estação meteorológica..... 55
6.3.4 Modelo Dimensional 55
6.3.5 Aplicação da Ferramenta OLAP 57
6.4 COMPARATIVO DAS TÉCNICAS.....59
6.5 CONSIDERAÇÕES SOBRE O CAPÍTULO 60
7 CONSIDERAÇÕES FINAIS..... 61
7.2 CONTRIBUIÇÕES.....62
7.3 TRABALHOS FUTUROS E MELHORIAS 63
8 REFERÊNCIAS 65
APÊNDICE A – CONFIGURAÇÃO MONDRIAN/PENTAHO..... 70
A.1 PRÉ-REQUISITO 70
A.2 INSTALAÇÃO.....70
A.3 CONFIGURAÇÃO DO BANCO DE DADOS 72
A.4 CRIAÇÃO DO SCHEMA PARA CUBO OLAP.....73
APÊNDICE B – CODIGOS SQL 77
B.1 INSERT NA TABELA FATO.....77
B.2 INSERT NA DIMENSÃO TEMPO 78

1 INTRODUÇÃO

1.1 CONSIDERAÇÕES INICIAIS E MOTIVAÇÃO

A Amazônia é conhecida por sua diversidade de flora e pela grande reserva de água doce que contêm os rios da região, além de abrigar um gigantesco reservatório natural de espécies da fauna brasileira, muitas ainda não conhecidas pelos cientistas, enquanto outras já estão em extinção devido à exploração indevida e ao contrabando.

Dentre essas espécies presentes na região amazônica, destaca-se um grupo de agentes polinizadores que são responsáveis por transferir grãos de pólen de uma flor para outra, e consequentemente contribuírem para o desenvolvimento de frutos e sementes. Alguns seres vivos pertencentes a esse grupo são: borboletas, besouros, morcegos, aves e abelhas (MILET-PINHEIRO e SCHLINDWEIN, 2008).

A polinização representa, atualmente, um fator de produção fundamental na condução de muitas culturas agrícolas ao redor do mundo e as abelhas destacam-se como os principais agentes polinizadores dos vegetais; em troca, os vegetais produzem substâncias adocicadas que atraem as abelhas, as quais levam, em seus pelos, o pólen dessa planta florífera. O pólen é importante para o desenvolvimento da colmeia, pois é a fonte principal de proteína das abelhas; assim, ao garantir o desenvolvimento da sua família, as abelhas também perpetuam a espécie vegetal (SOUZA, EVANGELISTA-RODRIGUES e PINTO, 2007).

Há ainda casos nos quais a existência de determinadas espécies de abelha estão condicionadas à presença de certas classes de fauna e vice-versa (MILET-PINHEIRO e SCHLINDWEIN, 2008). Além disso, as abelhas apresentam comportamento social bem elaborado em algumas espécies, o que tem gerado alguns estudos científicos nessa área (MELO, 2000).

Dentre os estudos envolvendo abelhas destacam-se aqueles que utilizam tecnologia de ponta para monitorar o *habitat* desses animais a fim de identificar comportamentos em diferentes momentos do dia, considerando variações de temperatura, luminosidade, umidade e até contabilizar a quantidade de abelhas de

uma colmeia. Para esse monitoramento, uma alternativa recente que vem sendo pesquisada é a utilização de Redes de Sensores sem Fio (RSSF) (CUGNASCA, 2008).

As RSSFs constituem-se em dispositivos que não utilizam fios como meio de comunicação, e que podem ser densamente distribuídos em uma região de interesse em se monitorar . Os nós sensores são os elementos principais desse tipo de rede, e possuem espaço limitado para armazenamento de dados; porém, são capazes de coletar dados de diferentes ambientes e de se comunicarem, visando transferir as informações a distâncias maiores do que o alcance de cada um (AKYILDIZ, SU, *et al.*, 2002). Existem diversas restrições quanto ao uso de RSSFs em relação ao seu consumo de energia, pois os nós são alimentados por baterias, que dependendo da utilização, a energia pode durar dias, semanas ou meses. Tal característica mostrou-se crítica em vários experimentos, principalmente aqueles localizados em áreas afastadas, nas quais a troca das baterias é mais difícil. Por isso, faz-se necessária a criação de alternativas para minimizar o consumo de energia (MIN, BHARDWAJ, *et al.*, 2001) (MINI, NATH e LOUREIRO, 2002), dentre elas destacando-se o uso de células solares para recarregar as baterias (MINAMI e MORITO, 2005).

A aplicabilidade das RSSFs é grande: sua utilização pode ir desde monitoração de aeronaves, com nós sensores acoplados às asas dos aviões, até a monitoração de complexos ecossistemas, com nós sensores instalados em coleiras de animais selvagens (ESTRIN, GOVINDAN e HEIDEMANN, 2000) (ESTRIN, GIROD, *et al.*, 2001). A monitoração de *habitat* e espécies animais ou vegetais com RSSFs apresenta-se como um campo de pesquisa com grande potencial de benefícios para as comunidades científicas e para a sociedade como um todo (MAINWARING, CULLER, *et al.*, 2002) (CUGNASCA, 2008) (RAMANATHAN, HARMON, *et al.*, 2006) (POLASTRE, 2003) (WERNER-ALLEN, JOHNSON, *et al.*, 2005) (WERNER-ALLEN, LORINCZ, *et al.*, 2006) (JUANG, OKI, *et al.*, 2002).

Juang *et al.* (2002) apresentaram a utilização de RSSF para observar o comportamento de zebras, cavalos selvagens e leões num centro de pesquisa no Quênia. O experimento durou cerca de um ano e monitorou uma área de milhares de quilômetros quadrados, monitorando os animais em situações como: na interação com outros animais, da mesma espécie ou não; ao pastarem; em movimento; e ao contato com espécies humanas.

Cugnasca (2008) abordou a utilização de redes de controle e RSSF para o estudo de abelhas. No trabalho foi utilizado um nó sensor para o monitoramento das colônias, que mede temperatura e umidade internas, contagem da entrada e saída de abelhas e possui um microfone para captura dos sons gerados pelas abelhas. Externamente as grandezas do ambiente eram observadas por uma estação meteorológica.

Com o auxílio da RSSF é possível monitorar diversas características dos ambientes nos quais os animais, inclusive as abelhas, estão presentes; porém, esses dados isolados ou simplesmente coletados ao longo do tempo podem dificultar a interpretação pelos usuários, e principalmente podem não contribuir para a rápida tomada de decisões, pois por característica, as RSSFs geram grandes quantidades de dados, mutáveis com frequência, não consolidados e às vezes, não sincronizados.

Seja um cenário hipotético onde há uma RSSF composta por vinte nós sensores espalhados num local de experimento coletando dados sobre temperatura, umidade realtiva do ar e luminosidade a cada dez segundos. Sabendo que cada coleta de dados por um nó sensor gera um registro, em cerca de uma hora, tem-se aproximadamente 7200 registros do experimento.

Esses dados coletados pela RSSF precisam ser processados e combinados para gerar informação. A informação é o agrupamento de dados, que relacionados e inseridos num contexto útil e com sentido bem definido, auxiliam a tomada de decisões. A interpretação das relações entre dados e informação gera o conhecimento.

Um exemplo de dado é 15°C. Aplicando um contexto para esse dado obtém-se uma informação: A temperatura máxima do dia em São Paulo foi 15°C. O conhecimento é gerado a partir da interpretação da informação, por exemplo: Se a máxima do dia em São Paulo foi 15°C é provável que amanhã chova.

O ser humano possui limites para lidar com informação em quantidade, seja pelo volume ou pela frequência. Além disso, a informação pode ter diferentes origens, características e qualidade.

Contudo as ferramentas de tecnologia da informação tendem a diminuir o esforço do ser humano em lidar com os dados, simplificando o processo de geração de informação e conhecimento, por meio da diminuição das dificuldades na manipulação de grandes volumes de dados.

No contexto do gerenciamento de RSSFs, algumas pesquisas foram desenvolvidas com o objetivo de extrair os dados das redes formadas por nós sensores como as de Yao e Gehrke (2002) e Madden *et al.* (2005). Ambas têm o foco em um elemento importante de uma RSSF: o nó sensor, incluindo a alteração no código-fonte do seu programa ou atribuição de responsabilidade diferente para este elemento.

Assim, percebe-se que para que os dados monitorados pelas RSSFs sejam aproveitados de maneira produtiva pelos interessados, os mesmos devem estar organizados em um repositório ou banco de dados, e possuir uma interface de acesso rápido e fácil, por meio da qual o usuário possa visualizar informações consolidadas e tenha possibilidade de tomar decisões estratégicas.

Tal descrição remete ao conceito de Depósito de Dados (do inglês, *Data Warehouse* ou simplesmente DW) que significa um conjunto de tecnologias de suporte à decisão utilizado por pessoas interessadas em tomar decisões rápidas e de maneira fácil (INMON, 1997). Os DWs inicialmente foram concebidos para consolidar dados corporativos e auxiliar na tomada de decisões de gestores de empresas (CHAUDHURI e DAYAL, 1997), pois quando bem projetados contribuem diretamente para a competição das empresas (INMON, 1997). Com o avanço dessa tecnologia sua aplicação também vem crescendo e ganhando espaço em diversas outras áreas.

A utilização de DW para monitoração de dados sobre animais é uma área em expansão, na qual se destacam alguns trabalhos (MARQUES, 2002) (TONI, 2000).

Em Toni (2000) é apresentado o uso de *Data Warehouse* em cooperativas agrícolas. O trabalho discute a necessidade do acesso eficiente às informações, focando na gestão dessas cooperativas. Destaca-se nessa pesquisa a possibilidade de aplicação do caso em outros tipos de cooperativas a fim de se obter resultados semelhantes ou melhores.

Já Marques (2002) apresenta uma aplicação de DW para analisar dados sobre melhoramento genético da raça bovina nelore, os conceitos de DW e *Data Mart* (subconjunto de um DW). A proposta é apresentada com ferramentas proprietárias, o que geralmente pode inviabilizar financeiramente sua utilização por outros setores ou ainda customizações, quando necessárias.

Mesmo com uma ampla variedade de aplicações de RSSF, faltam alternativas para o gerenciamento dos dados coletados por meio delas. Uma alternativa é o DW,

tecnologia que está crescendo muito rapidamente e com grande apelo no suporte a tomada de decisões estratégicas. Assim, justificando a realização de pesquisas baseadas nessa tecnologia, como a realizada neste trabalho.

Assim, este trabalho propõe a utilização de DW para gerenciamento de dados de RSSF que monitoram polinizadores, incluindo a criação de um modelo para extrair, transformar e carregar os dados das redes de nós sensores em um ambiente criado com auxílio de ferramentas com código aberto, que permitiram obter relatórios, gráficos e pesquisas customizadas sobre o experimento monitorado.

1.2 OBJETIVO

Este trabalho teve como objetivo a aplicação do conceito de *Data warehouse* para a agregação, gerenciamento e apresentação de dados coletados por meio de RSSFs que monitoram polinizadores. Considerou-se na proposição das soluções a integração de dados de outras fontes, como estações meteorológicas e redes LonWorks, visando conferir maior flexibilidade, possibilitando a expansão para experimentos que monitorem outros animais.

Além disso, buscaram-se outros objetivos específicos, como:

1. Investigar o processo de coleta de dados de experimentos com RSSF para transformá-los em dados a serem carregados em DW.
2. Modelar um sistema de informação focando no processo de coleta dos dados das RSSF, carga no DW e apresentação das informações consolidadas para os usuários.
3. Criar um protótipo funcional que permitisse validar a modelagem do sistema proposto.

1.3 CONTEÚDO E ORGANIZAÇÃO

O Capítulo 2 apresenta uma revisão sobre os fundamentos da tecnologia Redes de Sensores sem Fio, desde a sua criação até o estado da arte, destacando em particular uma RSSF destinada a pesquisas científicas, desenvolvida por uma empresa americana, e que foi utilizada nesse trabalho.

O Capítulo 3 apresenta uma revisão sobre os fundamentos dos DWs, a história, conceitos relacionados, destacando seu importante papel para tomada de decisões.

O Capítulo 4 destaca os trabalhos relacionados à monitoração de animais e *habitat* com RSSF, e em seguida apresentam-se algumas pesquisas que utilizam DW para gerenciar dados de animais, e por fim, destaca outras opções já existentes para gerenciamento de dados coletados por RSSF. O Capítulo é finalizado destacando o porquê da utilização de DW neste trabalho.

O Capítulo 5 apresenta e discute a arquitetura proposta para o protótipo funcional, tratando de maneira genérica os dados coletados por RSSFs e até mesmo os dados de outras fontes.

O Capítulo 6 apresenta e discute o protótipo funcional concebido para validar as pesquisas do presente trabalho, por meio do qual os dados de experimentos coletados por RSSFs são relacionados com dados de outras fontes e após serem tratados, são carregados em um DW, capaz de gerar análises complexas e flexíveis para os interessados.

Finalmente, o Capítulo 7 apresenta as considerações finais deste trabalho, destacando os benefícios da abordagem escolhida, as contribuições e possíveis melhorias, bem como trabalhos que poderiam ser considerados em continuidade a este.

O Apêndice A contém um guia para instalação e configuração do Mondrian/Pentaho, solução utilizada para criação do protótipo funcional.

2 REDES DE SENSORES SEM FIO

2.1 INTRODUÇÃO

As Redes de Sensores sem Fio (RSSF) têm sido viabilizadas pela rápida convergência de três tecnologias: microeletrônica, comunicação sem fio e micro sistemas eletromecânicos (SILVA, BRAGA, *et al.*, 2004). Uma RSSF é um recurso de sensoriamento de fenômenos, com arquitetura distribuída envolvendo o processamento e disseminação de dados coletados, e as informações processadas para um ou mais observadores. O potencial de observação e controle do mundo real permite que as RSSFs se apresentem como uma solução para diversas aplicações de monitoração e controle, tais como monitoração ambiental, gerenciamento de infraestrutura, monitoração e controle industrial, segurança pública e de ambientes em geral, áreas de desastres e risco para vidas humanas, transporte, medicina e controle militar (MEGUERDICHIAN, KOUSHANFAR, *et al.*, 2001).

Essas redes também podem ser utilizadas em aplicações de monitoração nas quais o custo de instalação e manutenção de cabos é um fator crítico. A sua instalação não exige infraestrutura (cabos de rede ou energia). Além disso, as RSSFs têm capacidade de processar os dados de forma distribuída, uma vez que cada nó sensor é equipado com processador e memória. Este aspecto distingue os elementos de uma RSSF da tecnologia de sensoriamento remoto tradicional, pois permite a percepção do ambiente de forma não intrusiva e com qualidade não disponível anteriormente. Esta é uma tecnologia relativamente recente, que permite integrar o mundo físico a outras redes, como a Internet (ESTRIN, GOVINDAN e HEIDEMANN, 2000).

Em função da área (remota ou inóspita) onde os nós sensores são depositados e do número de elementos que compõem a rede (centenas ou milhares), a substituição ou recarga manual das fontes de energia desses elementos pode não ser uma atividade viável. Assim, uma das maiores restrições à utilização de uma RSSF é o consumo de energia que ela apresentará (GANESAN, CERPAA, *et al.*, 2003).

2.2 NÓ SENSOR

Um nó sensor é um dispositivo eletrônico com capacidade de processamento, memória e interface de comunicação sem fio, além de possuir um ou mais sensores acoplados. Esse dispositivo de sensoriamento pode ser equipado com uma variedade de dispositivos sensores, tais como acústicos, sísmicos, infravermelhos, vídeo-câmeras, de calor, de temperatura e de pressão. Quanto à comunicação sem fio entre os sensores, ela pode pode ser ótica ou usar radiofreqüência, que é a forma mais comum (LOUREIRO, NOGUEIRA, *et al.*, 2003).

Esta capacidade de processamento se deve à presença de um microcontrolador no nó sensor, que possui memória interna, que algumas vezes é expandida externamente. A comunicação sem fio é geralmente realizada por meio de um transceptor de rádio (trasmissor/receptor). A alimentação do dispositivo advém de baterias. Os sensores são os responsáveis pela aquisição de dados de grandezas físicas ou eventos do mundo real, que quando processados podem ser transformandos em informações difundidas pela rede (AKYILDIZ, SU, *et al.*, 2002).

A arquitetura típica de um nó sensor é apresentada na Figura 1.

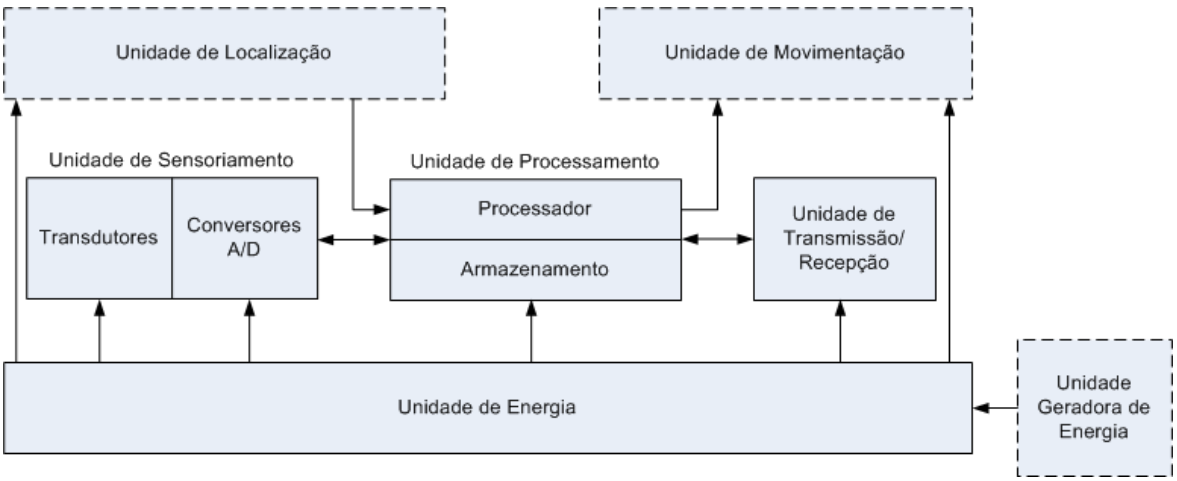


Figura 1. Arquitetura de um nó sensor. Adaptada de (AKYILDIZ, SU, *et al.*, 2002).

Os quatro elementos básicos que compõem a arquitetura básica de um nó sensor são:

1. Unidade de sensoriamento;
2. Unidade de processamento;
3. Unidade de transmissão/recepção;
4. Unidade de energia.

Além dos quatro elementos básicos, o nó sensor pode possuir outros elementos, dependendo da aplicação: unidade de localização, unidade de geração de energia e unidade de movimentação.

A unidade de sensoriamento é composta por duas subunidades, os transdutores e os conversores A/D. Os sinais analógicos produzidos pelos sensores nas regiões de monitoração são convertidos em sinais digitais e processados pela unidade de processamento. Esta, por sua vez, é geralmente associada a uma pequena unidade de armazenamento de dados, e fica responsável pelos procedimentos que fazem os sensores colaborarem entre si para realizar a tarefa de sensoriamento. A unidade de energia é composta por células de energia, que podem conter pilhas ou baterias, e ainda podem estar ligadas a uma unidade de geração de energia associada, como uma célula solar que alimente as células de energia, prolongando o tempo de monitoração sem interrupções para troca de baterias. A unidade de transmissão/recepção é responsável pela comunicação na camada física da rede (AKYILDIZ, SU, *et al.*, 2002).

Enquanto que a unidade de localização é um elemento opcional num nó sensor, sua utilização é muito útil, pois a maioria das técnicas de roteamento conhecidas necessita da localização com alta exatidão. Então é comum que um nó-sensor tenha uma unidade de localização. Algumas vezes uma unidade de movimentação é necessária para a realização de tarefas específicas, como a medição da aceleração nos eixo x e y (AKYILDIZ, SU, *et al.*, 2002).

Normalmente todas estas unidades devem possuir pequenas dimensões (por exemplo, do tamanho de uma caixa de fósforos), ou até mesmo de volumes menores, próximos a um centímetro cúbico.

Além do tamanho, há outros requisitos para os nós sensores (AKYILDIZ, SU, *et al.*, 2002):

- apresentar baixo consumo de energia;
- operação em altas densidades volumétricas, grandes campos ou lagos, por exemplo;
- possuir baixo custo de produção e até mesmo ser descartável;
- operar autonomamente;
- adaptar-se ao ambiente de utilização.

2.3 *GATEWAY*

No contexto considerado, o *gateway* é um equipamento responsável por fazer a interface entre uma RSSF e um computador, viabilizando que os dados coletados sejam armazenados em bancos de dados, para serem posteriormente tratados por programas específicos e visualizados na tela de um computador.

Antes dos dados serem enviados ao *gateway* de uma RSSF, eles trafegam por diversos nós sensores da rede, estando suscetíveis a perdas ou duplicidade de dados, devido à sobreposição de áreas de monitoramento e falhas na comunicação entre os nós sensores (NAKAMURA e LOUREIRO, 2008).

A Figura 2 apresenta a visão geral de um RSSF, representando: nós sensores e um *gateway*.

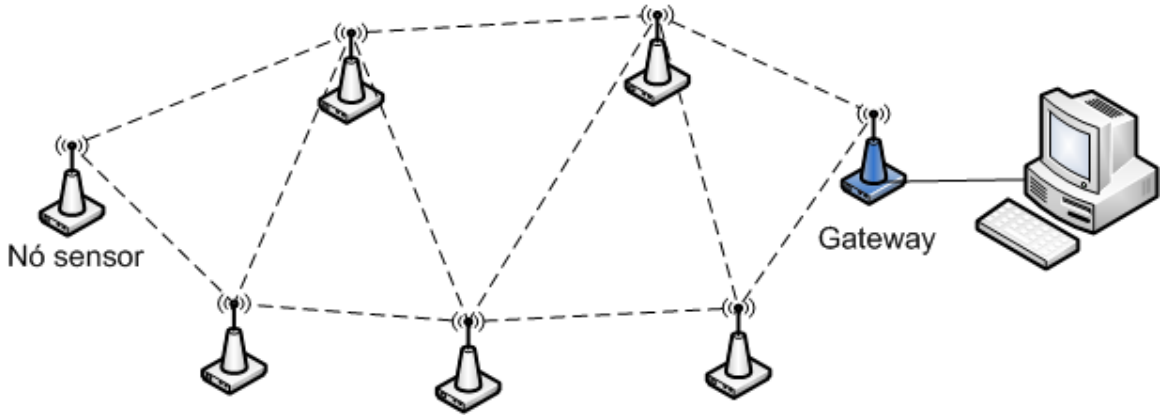


Figura 2. Exemplo de um RSSF.

2.4 SISTEMA OPERACIONAL TINYOS

O TinyOS é um sistema operacional de código aberto, utilizado nos nós sensores de uma RSSF, que possuem recursos limitados (memória e capacidade de processamento).

A arquitetura do TinyOS é baseada em componentes, sendo que as bibliotecas disponíveis no sistema operacional incluem protocolos de rede, serviços distribuídos, *drivers* para sensores e ferramentas para aquisição de dados, as quais podem ser úteis para criação de aplicações para funcionalidades específicas (TINYOS, 2010).

2.5 PROTOCOLOS DE COMUNICAÇÃO

As RSSF têm características que as diferem bastante das redes mais comuns, sendo uma delas a centralização em dados, diferentemente das redes tradicionais, que são centradas em endereço. Assim, um nó difunde dados baseadas em atributos. Além disso, os nós sensores devem atender a requisitos específicos da aplicação. Muito comumente os nós focam-se em apenas em um atributo, ou um pequeno conjunto de atributos, necessitando, assim, de processamento no interior da rede. As restrições impostas à RSSF implicam em uma série de requisitos especiais para os protocolos de comunicação e gerenciamento da rede, que devem ter capacidade de auto-organização.

2.5.1 ZigBee

O padrão ZigBee (BARONTI, PILLAI, *et al.*, 2007) estabelece um conjunto de especificações (associado a uma camada de protocolos) para a comunicação sem fio em sistemas de monitoramento e controle. Estes sistemas costumam apresentar como principais características a exigência de pouca largura de banda de

comunicação (as mensagens tipicamente são de pequeno tamanho), o consumo de energia deve ser baixo, assim como a latência, considerando o grande número de dispositivos que podem constituir uma rede.

O padrão IEEE 802.15.4 contém as especificações para as camadas mais baixas da pilha de protocolos ZigBee (física e de acesso ao meio).

As características do padrão ZigBee/IEEE 802.15.4 (KINNEY, 2003):

- a camada Física pode operar em duas faixas de frequência (2.4 GHz e 868/915 MHz);
- a taxa de transmissão de dados pode ser de 250 kbps (2.4 GHz), 40 kbps (915 MHz) e 20 kbps (868 MHz);
- o acesso ao meio físico segue o protocolo CSMA-CA;
- o endereçamento utiliza 64 bits (endereço que segue o padrão IEEE);
- dois tipos de dispositivos físicos são previstos, o *Full Function Device* (FFD) e o *Reduced Function Device* (RFD). Cada um desses dispositivos pode estar ativo (transmitir/receber) ou inativo;
- admite as topologias de rede (estrela, ponto a ponto e malha);
- o alcance é de 50 m.

2.6 PROTOCOLO DE ROTEAMENTO

Em Heinzelman; Chandrakasan e Balakrishnan (2002), Intanagonwiwat; Govindan e Estrin (2000) e Shah, *et al.* (2003) são apresentados alguns tipos de protocolos de roteamento, os quais podem ser adotados dependendo do tipo de aplicação a ser implementada. O presente trabalho abordará o protocolo Xmesh (XMESH, 2009), o qual vem embarcado na tecnologia de RSSF da rede experimental utilizada, apresentada na seção 2.5.

O Xmesh, baseado nas especificações do padrão ZigBee, permite redes que se auto-organizam, *multi-hop* e *ad-hoc*. Uma rede XMesh é formada por nós sensores, chamados de *motes*, os quais têm a capacidade de encaminhar seus dados por meio da rede, até atingir a estação base, o *gateway*, de forma que esta possa transmitir esses dados ao computador. O Xmesh fornece dois benefícios

principais, que são a melhora na cobertura do rádio e melhora na confiabilidade, permitindo que as redes se auto-organizem e sejam tolerantes a falhas (XMESH, 2009).

2.7 CARACTERÍSTICAS

Dentre as características das RSSFs destacam-se as seguintes (AKYILDIZ, SU, *et al.*, 2002):

- **Eficiência de energia**: os nós sensores possuem severas restrições de energia, pois são geralmente alimentados por pequenas baterias ou pilhas de curta duração, e nem sempre é viável a sua troca ou recarga, dependendo do ambiente de monitoração. Assim, cada sensor deve ser o mais eficiente possível. Tal restrição mostra-se crítica em algumas aplicações. Minami e Morito (2005), Mini; Nath e Loureiro (2002) e MIN, *et al.* (2001).
- **Baixo custo**: os nós sensores devem apresentar custo reduzido, pois determinados experimentos exigem que centenas ou milhares deles sejam espalhados sobre a região de interesse em se monitorar; outras aplicações não preveem o reaproveitamento dos nós, que são descartados após o uso.
- **Comunicação Sem Fio**: é uma premissa dessa tecnologia que independe de infraestrutura de cabos.
- **Facilidade de Programação**: deve ser possível reprogramar os nós sensores para que, por exemplo, obtenham-se dados em períodos diferentes ou ainda para que se altere o modo de comunicação e troca de dados com os outros nós sensores.
- **Expansibilidade:** o hardware dos nós pode prever a incorporação de outros tipos de sensores, além dos já existentes, para atender a necessidades de aplicações específicas.
- **Tamanho:** as dimensões reduzidas dos nós sensores são uma característica importante em diversas aplicações, como no caso da monitoração de animais, que exigem facilidade de instalação e minimização do incômodo.

2.8 UM EXEMPLO DE REDES DE SENSORES SEM FIO

Dentre a variedade de nós sensores para RSSF destacam-se alguns tipos desenvolvidos pela Universidade Berkeley, em conjunto com os Laboratórios Intel Research e comercializados pela Crossbow Technology Inc. (CROSSBOW, 2010), que em 2010 transferiu a sua divisão de RSSF para a empresa Memsic (MEMSIC, 2010). Eles costumam ser utilizados em pesquisas acadêmicas, como a presente. Foram utilizados os nós sensores pertencentes à família MicaZ, que requerem como estação base o *gateway* USB MIB 520 (MEMSIC, 2010).

A placa de sensores de um nó utilizada foi a modelo MTS 400, que possui cinco tipos de sensores: umidade relativa do ar, temperatura, pressão barométrica, luminosidade e aceleração em duas dimensões ortogonais (MEMSIC, 2010). Ela é conectada a uma placa com um transceptor de radiofreqüência baseado no padrão IEEE 802.15.4, a MPR 2400, que permite uma taxa de transmissão de dados de até 250 kbps (Figura 3). A sua banda de operação é de 2.400 MHz a 2.483,5 MHz, utilizando o rádio Chipcon CC2420. O processamento do nó fica a cargo do microcontrolador Atmega128L, que possui memória *flash* interna de 128 Kbytes, memória *flash* externa serial de 512 Kbytes e EEPROM 4 Kbytes para dados de configuração do nó. Também possui um conector de entrada/saída de 51 pinos, que permite a interface com diversos tipos de placa de sensores, como a MTS 400 citada. A alimentação do conjunto é feita por duas pilhas AA, residentes no módulo MPR 2400, com um intervalo de tensão de operação de 2,7 VDC a 3,3 VDC (MEMSIC, 2010).

Figura 3. Placa MPR2400. Extraído de (MEMSIC, 2010).

Um exemplo de *gateway* para RSSF da Memsic é apresentado na Figura 4, o modelo MIB520, que possui uma interface padrão USB de conexão com um computador, e um conector para acoplamento da placa com um transceptor de radiofreqüência MPR 2400, o mesmo tipo utilizado nos nós sensores. Esse mesmo conector é utilizado para a programação de cada nó. Com o transceptor de radiofreqüência o MIB520 se transforma em um *gateway* de uma RSSF (MEMSIC, 2010).

Figura 4. MIB 520 – *Gateway* USB. Extraído de (MEMSIC, 2010).

2.9 CONSIDERAÇÕES SOBRE O CAPÍTULO

Neste capítulo foram apresentados algumas características básicas das RSSFs e formas típicas de sua utilização. Alguns detalhes da RSSF para pesquisas disponível foram incluídos, por serem de interesse à modelagem do sistema proposto nesse trabalho.

3 DATA WAREHOUSE

3.1 INTRODUÇÃO

O *Data Warehouse* (DW) é uma coleção de dados orientados por assuntos, integrados, variáveis com o tempo e não voláteis, para dar suporte ao processo gerencial de tomada de decisões (INMON, 1997).

Suas principais características são:

- **Orientado a assunto**: enquanto que os sistemas tradicionais são sistemas transacionais e voltados para os processos de uma empresa e aplicações específicas, o DW foca sua modelagem em assuntos específicos, geralmente segmentos particulares dentro de uma empresa. Por exemplo, uma empresa seguradora de automóveis, pode organizar seus dados por consumidor, franquia e prêmio, conforme representa a Figura 5, e assim identificar comportamentos e tendências dentro de cada um desses assuntos.

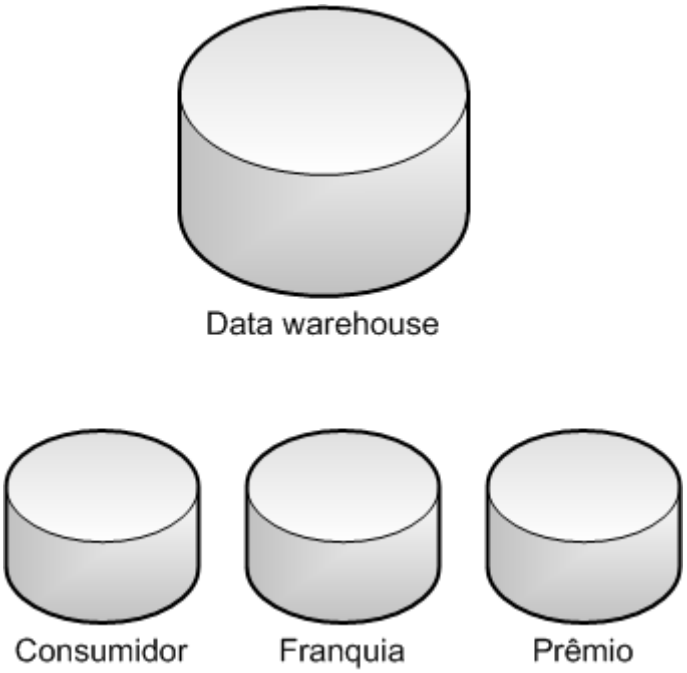


Figura 5. *Data Warehouse* orientado a assuntos.

- **Integrado**: quando os dados estão armazenados em apenas uma base de dados, geralmente já estão organizados e padronizados de acordo com o padrão adotado na criação do mesmo. Isso se refere à definição de apresentação para data e hora, valores monetários e casas decimais. Tais definições podem variar de acordo com cada projeto de banco de dados. Alguns bancos de dados adotam 'M' ou 'F' para representa o sexo em campos de uma tabela, enquanto que outros bancos utilizam o padrão '0' (masculino) e '1' (feminino). Outro exemplo é o sistema de medida alguns bancos de dados podem ter sido projetados para utilizar metros enquanto outros centímetros. No DW todos esses dados devem estar padronizados e estas padronizações fornecem a forma pela qual as informações de diversos sistemas e fontes de dados podem ser relacionadas, conferindo-lhe, assim, propriedade de ser integrado. A Figura 6 apresenta um exemplo de conversão de dados sobre temperatura obtida de diferentes bancos de dados e que necessitam ser integrados no DW de maneira padronizada.

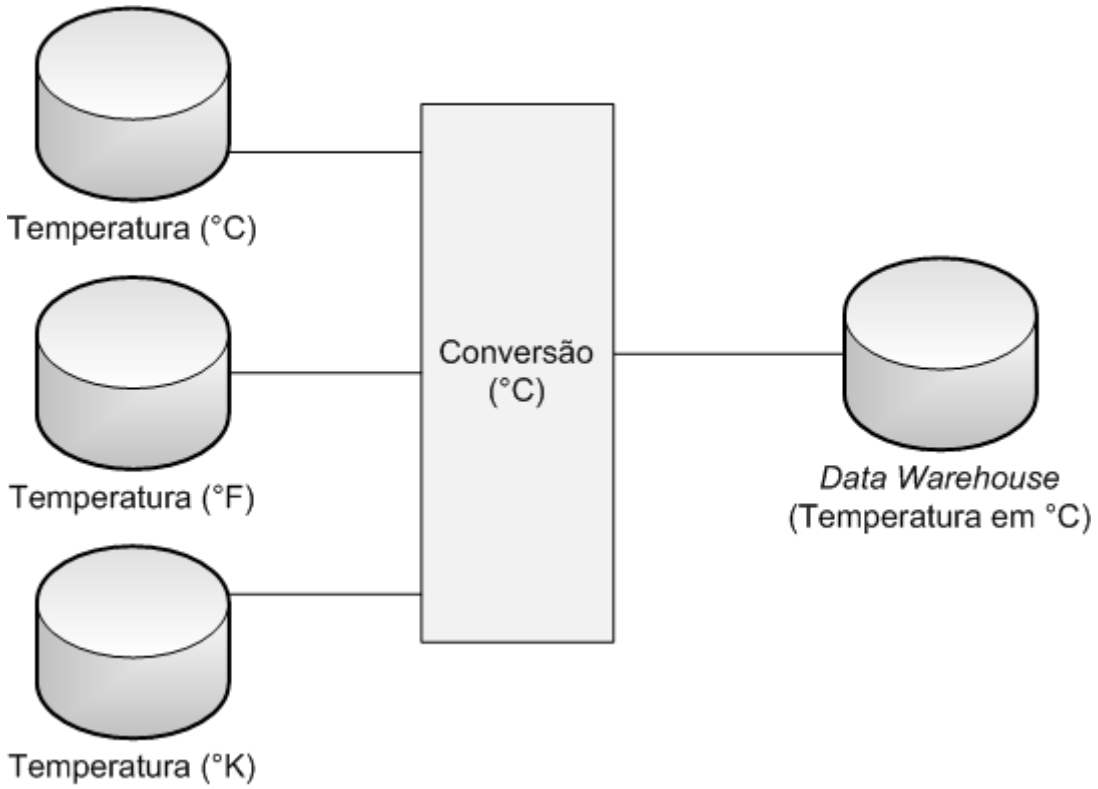


Figura 6. Dados sobre temperatura sendo convertidos antes de serem colocados no DW.

- **Não volátil**: por definição os dados armazenados em DWs são estáveis e não são mais removidos. Atualizações são raras. Assim as principais operações são cargas e consultas. Isso permite uma melhor visão geral do negócio que é gerenciado por meio do DW. Por exemplo, supondo um DW que contenha informações sobre Clientes de uma empresa (usados para uma ação de marketing), após alguns anos o cliente procura a empresa para informar que o seu endereço de correspondência mudou. Tal operação deve ser realizada no DW conforme mostra a Tabela 1. Observa-se que o registro prévio não foi atualizado, mas um novo registro foi criado. Dessa forma os relatórios que solicitarem os clientes até a data de 10 de Abril de 2011 trarão os produtos vendidos dessa Empresa como sendo resultados das campanhas das equipes de marketing da região Centro e posteriormente a esta data as vendas serão atribuídas às equipes responsáveis pela região Sul.

Tabela 1. Exemplo de não volatilidade de dados em DW.

Nome Cliente	Endereço	Numero	Região	Data do Registro
João da Silva	Rua da Rosas	10	Centro	1/3/2005
João da Silva	Rua X de Março	32	Sul	10/4/2010

- **Variável com o tempo**: para realizar comparações e análises de tendências o DW armazena dados históricos, que podem variar 1, 5, 10 anos ou mais. Estes dados não são atualizados.

O DW fornece acesso a dados corporativos ou organizacionais. Esses dados são consistentes, podendo ser separados e combinados usando-se qualquer medição possível no negócio. Um DW não consiste apenas em dados, mas em um conjunto de ferramentas para consultar, analisar e apresentar informações, um local onde se publicam dados confiáveis, que podem ajudar na tomada de decisões estratégicas por gestores de empresas ou de negócios (KIMBALL, 1997).

3.2 CARACTERÍSTICAS

Enquanto os bancos de dados tradicionais são transacionais (relacional, orientado a objetos, de rede, ou hierárquico), os DWs tem a característica distinta, pois são direcionados principalmente para aplicações de apoio às decisões.

Os DWs são projetados para a recuperação de dados, não para o processamento rotineiro de transações, dando suporte a vários tipos de aplicações, ferramentas *OLAP*, *DSS* e *Data Mining*, descritas a seguir.

3.3 *OLAP*

O *Online Analytical Processing* – Processamento Analítico On-Line (OLAP) é um termo utilizado para descrever a análise de dados a partir do DW. Nas mãos de especialistas, as ferramentas OLAP empregam as capacidades de computação distribuída para análises que requerem maior capacidade de armazenamento e poder de processamento do que podem oferecer os computadores tradicionais.

As ferramentas OLAP têm como função a navegação nos dados de um DW, possuindo uma estrutura adequada tanto para as pesquisas como para a apresentação das informações.

Com elas é possível navegar entre diferentes granularidades de um DW. Através de um processo denominado *Drill* o usuário pode aumentar (*Drill down*) ou diminuir (*Drill up*) o nível de detalhamento dos dados. Por exemplo, um relatório pode estar consolidado por Estado. Com o *Drill down* os dados passam a ser apresentados por cidades, bairros e assim sucessivamente até o menor nível possível. O processo contrário *Drill up* faz com que os dados sejam consolidados em níveis superiores.

Outra possibilidade apresentada pela maioria das ferramentas de navegação OLAP é a de rearranjar colunas e linhas. É possível trocar a ordem de colunas e linhas bem como suprimi-las ou exibir aquelas que estejam ocultadas na visualização dos dados.

A possibilidade de manipular dados e formas de apresentação de maneira rápida é um dos pontos fortes de um DW. Essa característica faz com que os relatórios sejam mais comumente utilizados em tela e não na forma impressa. Quem necessitar dos dados pode, então, navegar nas informações por diversas maneiras, e ao final pode imprimir e até mesmo salvar aquela visão para uma futura consulta.

3.4 *DSS*

Os *Decision Suport Systems* – Sistemas de Apoio à Decisão (DSS), também conhecidos como *Executive Information Systems* – Sistemas de Informações executivas (EIS), dão apoio aos tomadores de decisão, líderes de uma organização com dados de mais alto nível para decisões complexas e importantes.

3.5 *DATA MINING*

O termo *Data Mining* - Garimpagem de Dados (DM) refere-se à mineração ou descoberta de novas informações em função de padrões ou regras em grandes quantidades de dados. O DM é orientado a descobertas, fornece percepções dos dados coorporativos que não podem ser obtidos com o OLAP, descobrindo padrões e relacionamentos ocultos em grandes bancos de dados e inferindo regras a partir deles para prever comportamentos futuros. Esses modelos e regras podem então ser utilizados para guiar o processo de decisão e prever os efeitos.

3.6 *DATA MART*

Segundo Laudon e Laudon (2007), empresas podem montar DWs de âmbito empresarial, nos quais um armazém central de dados atende toda a organização, ou podem criar armazéns menores, descentralizados, conhecidos como *Data Mart*

(DM). DM é um subconjunto de um DW, no qual uma porção resumida ou altamente focalizada dos dados da organização é colocada em um banco de dados separado destinado a uma população específica de usuários.

Sendo assim, os DMs podem ser orientados aos departamentos de uma empresa, como por exemplo, a área comercial, financeira, marketing e logística construindo-se um DM para cada uma.

De maneira geral, os DMs podem surgir de duas maneiras:

- **Top-down**: é quando os DM surgem após a criação do DW, ou seja, o DW é concebido e em seguida é dividido em áreas menores gerando assim pequenos bancos orientados por assuntos aos departamentos. Esta é a estratégia sugerida por Bill Inmon (INMON, 1997).
- **Botton-up**: é quando a situação é inversa. A empresa, por estratégia sua, prefere primeiro criar um banco de dados para somente uma área. Com isso os custos são bem inferiores de um projeto de DW completo. A partir da visualização dos primeiros resultados, as parte outras áreas passa a ser sucessivamente contempladas, até resultar em um DW. Está é a estratégia sugerida por Ralph Kimball (KIMBALL, 1997).

3.7 MODELAGEM DIMENSIONAL

Para discutir os DWs e distingui-los dos bancos de dados transacionais, exige-se um modelo de dados apropriado: o modelo de dados dimensional. Contrastando com os bancos de dados tradicionais, que proporcionam acesso a banco de dados disjuntos e geralmente heterogêneos, um DW é frequentemente um armazém de dados integrados oriundos de fontes múltiplas, processados para armazenamento em um modelo dimensional (ELMASRI e NAVATHE, 2005)

Os modelos dimensionais utilizam as relações inerentes aos dados para gerar dados em matrizes multidimensionais chamadas cubos de dados, que podem ter duas, três ou quantas dimensões forem possíveis ou aplicáveis.

A modelagem dimensional proporciona um ganho de tempo na consulta, uma melhor organização do sistema e principalmente a sua utilização de forma intuitiva pelo usuário.

O modelo dimensional é baseado em três elementos:

- Tabelas Fatos;
- Tabelas Dimensões;
- Medidas.

Um sistema concebido a partir da modelagem dimensional é composto de uma ou mais Tabelas Fatos que possuem as chaves estrangeiras das demais entidades relacionadas ao contexto, além de conter valores numéricos do negócio.

Cada medição é obtida por meio da intersecção das dimensões. As dimensões são tabelas que conterão os atributos textuais do negócio. A Figura 7 apresenta um exemplo de modelagem dimensional.

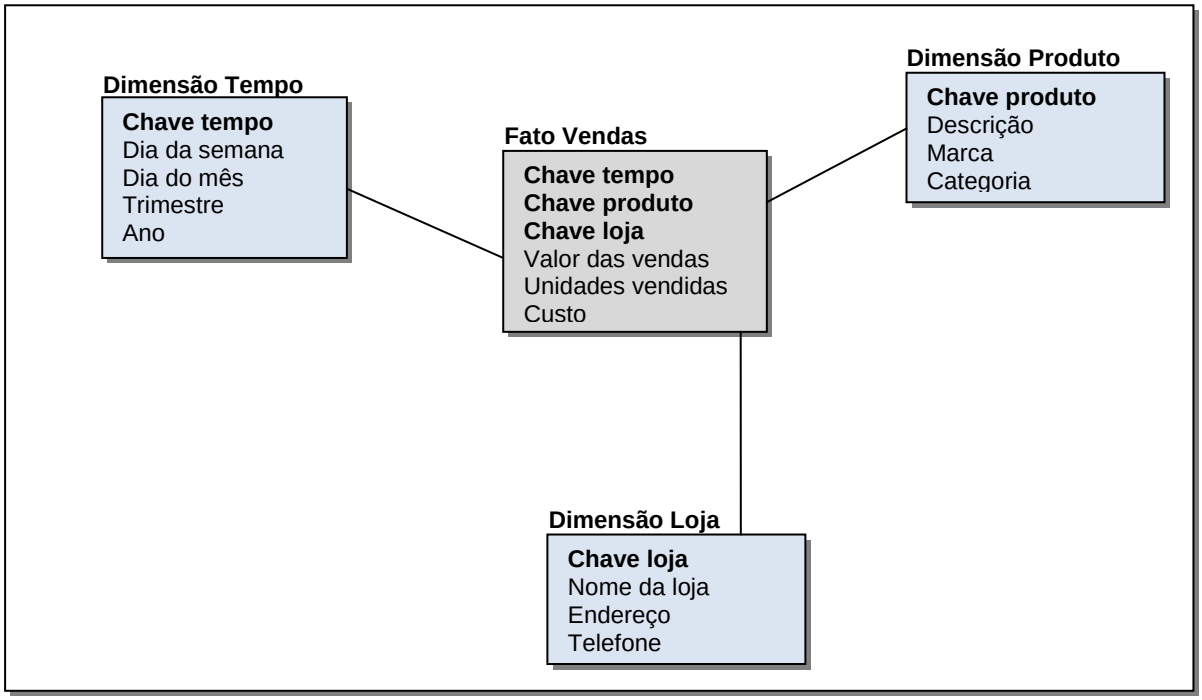


Figura 7. Exemplo de modelo dimensional com três dimensões e uma tabela fato.

A Tabela Fato guarda os dados do negócio, ou seja, todas as informações decorrentes do andamento do negócio que não são conhecidas anteriormente ao DW. Cada fato pode representar uma determinada transação, visão ou evento do negócio ocorrido em um determinado contexto obtido na intersecção das dimensões. Podem-se citar como exemplos, a venda de automóveis, o aluguel de apartamentos e a compra de imóveis.

A dimensão se refere ao contexto no qual um determinado fato ocorreu, tais como períodos de tempo, produtos, mercados, clientes e fornecedores e elementos

que possam descrever o contexto de um determinado fato. Por exemplo, o período (quando ocorreu a venda) e o modelo do carro.

Medidas são atributos que quantificam um determinado fato, representando a desempenho de um indicador em relação às dimensões que participam do fato. O contexto de uma medida é determinado em função das dimensões que participam do fato.

3.8 ESQUEMAS DIMENSIONAIS

São dois os esquemas dimensionais utilizados em DW: o esquema estrela (*star schema*) e o esquema floco de neve (*snow flake schema*). O primeiro consiste em uma Tabela Fato relacionando-se diretamente com todas as Tabelas Dimensões (ELMASRI e NAVATHE, 2005). Este modelo é chamado estrela porque a tabela fato fica ao centro e é cercada das tabelas dimensionais, conforme ilustra a Figura 8.

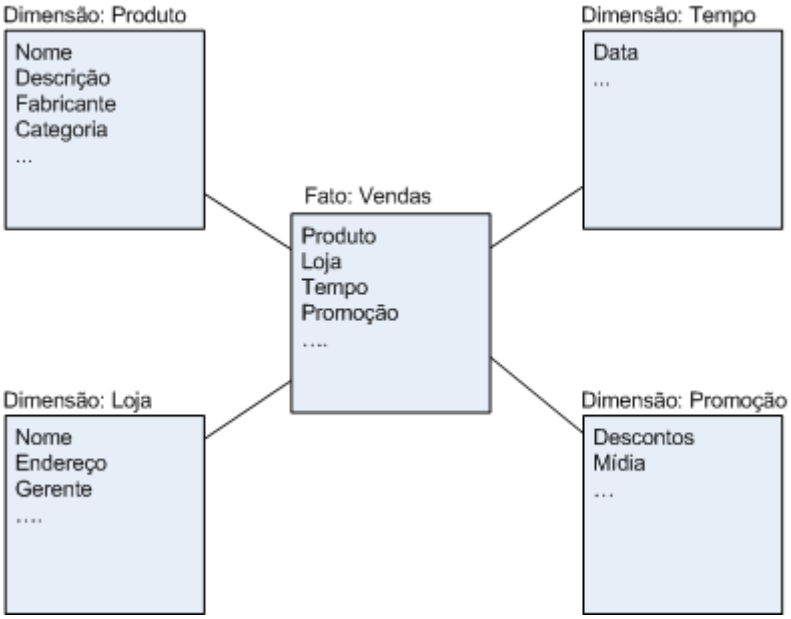


Figura 8. Esquema estrela (*star schema*).

O esquema floco de neve é uma variação do esquema estrela, no qual as Tabelas Dimensões são organizadas em uma hierarquia por meio da normalização.

Assim a Tabela Fato não se relaciona diretamente com todas as dimensões (ELMASRI e NAVATHE, 2005), conforme ilustra a Figura 9.

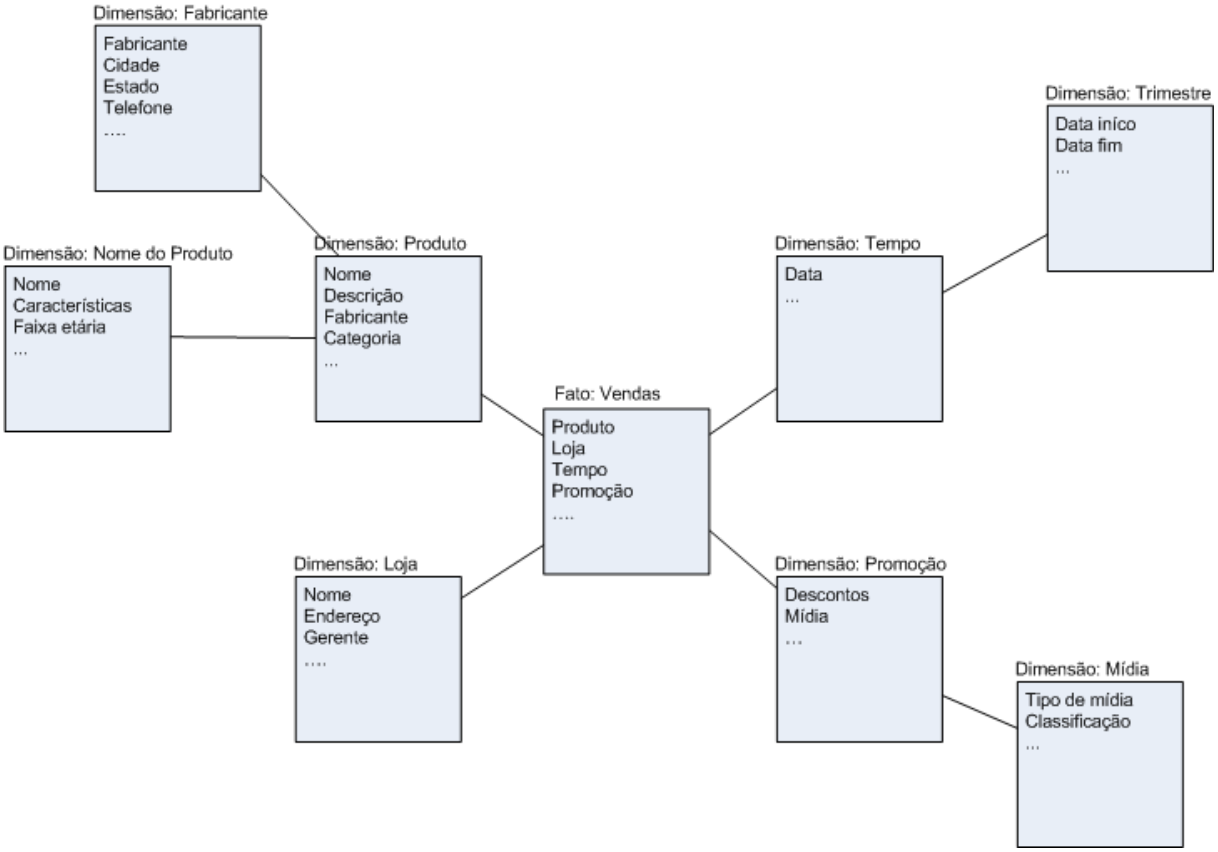


Figura 9. Esquema floco de neve (*snow flake schema*).

Em resumo, o esquema floco de neve pode reduzir o espaço de armazenamento dos dados dimensionais, porém acrescenta outras tabelas ao modelo, deixando-o mais complexo.

Já o esquema estrela é mais simples para modelar e para navegar, porém pode desperdiçar espaço repetindo informações ao longo de toda a tabela. Por exemplo, o atributo Fabricante de um produto não precisava ser repetido sempre que a dimensão Produto fosse utilizada.

3.9 ETL

O processo de *Extract Transform Load* – Extrair Transformar Carregar (ETL) é uma das fases mais críticas na concepção de um DW e também no momento de suas atualizações, pois envolve processamento dos dados.

Tal processo é ilustrado na Figura 10 e geralmente ocorre em horários de pouco ou nenhum acesso ao DW, geralmente pela madrugada, podendo ser uma operação lenta, que consume horas, dependendo do volume de dados a ser inserido.

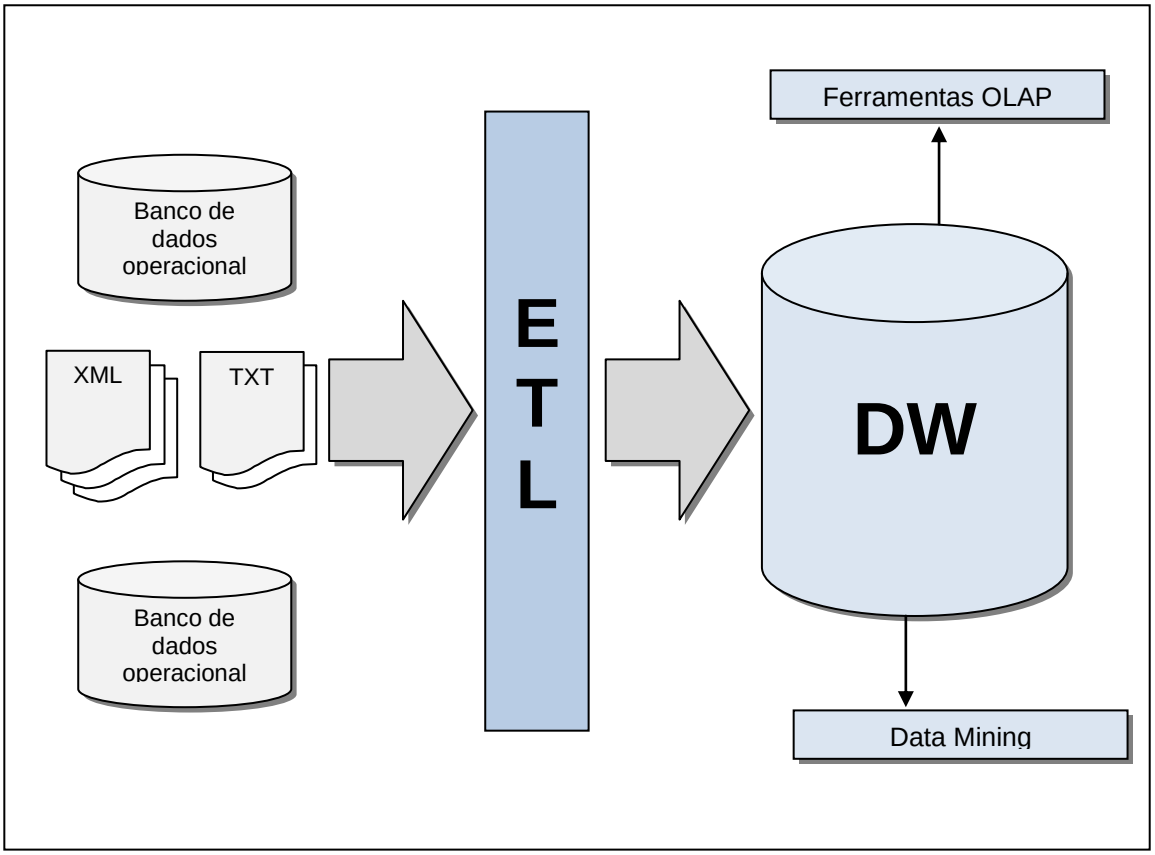


Figura 10. Processo ETL em um *Data Warehouse.*

O processo ETL ocorre em três etapas:

1) **Extração**: a primeira etapa consiste em definir as fontes de dados que serão consultadas. As origens podem ser banco de dados, arquivos XML, planilhas de texto, arquivos TXT e qualquer outra estrutura que armazene dados minimamente organizados.
2) **Transformação**: tal etapa representa a padronização dos dados de acordo com o tipo e representação desejada no DW. Por exemplo, o campo sexo pode estar representado de diferentes maneiras em várias fontes de dados (M e F; ou Masculino e Feminino; e até 0 ou 1), e para isso é importante que os mesmos sejam uniformizados antes de irem para o DW, e assim possam refletir compatibilidade e integração.
3) **Carga**: é última etapa e consiste na inserção dos dados, propriamente dita no DW.

3.10 MATRIZ DE BARRAMENTO

Matriz de barramento é o nome utilizado para representação dos processos de negócios (Fato) *versus* as variáveis relacionadas (Dimensões) no modelo dimensional. É a ferramenta utilizada para criar, documentar e comunicar arquitetura do DW. A Figura 11 representa um exemplo de matriz de barramento, na qual as células da matriz são marcadas caso uma determinada dimensão faça parte de uma determinada área de negócios (KIMBALL e ROSS, 2002).

	Dimensões				
Processos de Negócios (Tabelas Fato)	Produto	Loja	Tempo	Promoção	Fabricante
Vendas	x	x	x	x	x
Importação	x		x		x
Exportação	x	x			x

Figura 11. Exemplo de matriz de barramento.

Além disso, a representação por matriz permite flexibilidade na adição de novos processos, ares de negócios e dimensões, e assim integrá-las a já existente matriz.

3.11 CONSIDERAÇÕES SOBRE O CAPÍTULO

Neste capítulo foram apresentados os fundamentos do DW, incluindo as etapas para concepção e o processo de visualização dos dados. Eles são de interesse para a modelagem do sistema proposto nesse trabalho, principalmente o processo ETL.

4 TRABALHOS RELACIONADOS

4.1 MONITORAÇÕES DE ANIMAIS E *HABITAT* COM RSSF

A monitoração de *habitat* e espécies animais ou vegetais com RSSFs apresenta-se como um campo de pesquisa com grande potencial de benefícios para as comunidades científicas e para a sociedade como um todo (MAINWARING, CULLER, *et al.*, 2002), (CUGNASCA, 2008), (RAMANATHAN, HARMON, *et al.*, 2006), (POLASTRE, 2003), (WERNER-ALLEN, LORINCZ, *et al.*, 2006), (WERNER-ALLEN, JOHNSON, *et al.*, 2005),(JUANG, OKI, *et al.*, 2002).

Mainwaring e Culler (2002) apresentam uma arquitetura para monitoração de *habitats* com RSSFs ligadas a estações de testes localizadas na Ilha de Great Duck, Maine, EUA. Tal solução se destaca pelo baixo impacto gerado ao ambiente monitorado e pela economia dos gastos com deslocamento de pessoas aos experimentos para coleta de dados, visto que o sistema possui gerenciamento remoto por meio da Internet. Também se ressaltam os requisitos identificados e considerados para concepção da arquitetura, como: longevidades da RSSF, gerenciamento remoto da rede e armazenamento dos dados. Nesse trabalho foi utilizado o banco de dados PostgreSQL, seguindo o modelo relacional.

Werner-Allen et al. (2005) e Werner-Allen et al. (2006) apresentam experimentos de implantação de RSSF em vulcões. O trabalho foi realizado em parceria com estudiosos em vulcões e com Instituto de Geofísica do Equador, onde o experimento foi realizado. Os autores destacam uma limitação já conhecida dos nós de RSSFs, que é a baixa largura de banda de seus rádios. Citam como exemplo o dispositivo de rádio ‘Chipcon’ CC2420, que possui taxa de 30 kbps. Entretanto, *overheads* gerados pelo roteamento e quebra de pacotes podem diminuir a taxa para 10 kbps. Logo, os nós passam a adquirir dados mais rapidamente do que podem transmiti-los. Quanto à transmissão dos dados, os autores ressaltam que extrair alta fidelidade das RSSF é um desafio, pois os links de rádio são fracos e os osciladores a cristal têm baixa estabilidade, causando variações nas bases de tempo por toda a rede. Para o controle dos dados destaca-se a criação de aplicação Java

capaz de monitorar o experimento e definir parâmetros, como intervalos de amostragens e detecção de eventos. O computador que recebe os dados é responsável por armazenar as informações referentes aos pacotes transmitidos. Além disso, a aplicação apresenta uma tabela resumindo o estado da rede, baseada nas mensagens de estado enviadas periodicamente pelos nós.

As RSSFs também trouxeram grandes avanços para a agricultura (WARK, CORKE, *et al.*, 2007), (BAGGIO, 2005), (CAMILLI, 2004), (CAMILLI, CUGNASCA, *et al.*, 2007) e (GONDA, CUGNASCA e CAMILLI, 2005).

Gonda, Cugnasca e Camilli (2005) apresentam uma proposta de utilização para RSSF na agricultura de precisão (monitorando temperatura e umidade relativa do ar) no controle em casas de vegetação. O trabalho discute a organização dos sensores no cenário de uma casa de vegetação e na monitoração de uma cultura, destacando a importância da comunicação entre a RSSF e a estação base, preferencialmente por meio de comunicação sem fio. No ambiente descrito é relevante que os nós sensores estejam protegidos para resguardá-los de contato com água e insumos.

Juang *et al.* (2002) apresentam a utilização de RSSF para observar o comportamento de zebras, cavalos selvagens e leões num centro de pesquisa no Quênia. O experimento durou cerca de um ano e monitorou uma área de milhares de quilômetros quadrados. O principal objetivo da pesquisa foi monitorar os animais em situações como: interação com outros animais, da mesma espécie ou não; ao pastarem; em movimento; e ao contato com espécies humanas.

Os animais receberam nós sensores e um aparelho GPS, este último para obter a localização e a velocidade de movimento do animal. Os sensores coletavam os dados a cada três minutos e os transmitiam por meio da rede até uma estação, na qual temporária onde posteriormente um veículo coletava periodicamente os dados para armazenamento e verificação.

Cugnasca (2008) aborda a utilização de redes de controle e RSSF para o estudo de abelhas. No trabalho é utilizado um nó sensor para o monitoramento das colônias, que mede temperatura e umidade internas, contagem da entrada e saída de abelhas, controle da temperatura da colmeia. Externamente as grandezas do ambiente eram observadas por uma estação meteorológica.

4.2 *DATA WAREHOUSES* APLICADOS A MONITORAÇÃO DE ANIMAIS

A utilização de DW para monitoração de dados sobre animais é uma área em expansão, na qual se destacam alguns trabalhos.

Em Toni (2000) é apresentado o uso de DM em cooperativas agrícolas. O trabalho discute a necessidade do acesso eficiente às informações, focando na gestão dessas cooperativas. Destaca-se nessa pesquisa a possibilidade de aplicação do caso em outros tipos de cooperativas, a fim de obterem-se resultados semelhantes ou melhores.

Já Marques (2002) apresenta uma aplicação de DW para analisar dados sobre melhoramento genético da raça bovina nelore, o texto expõe amplamente os conceitos de DW e DM. A proposta é apresentada com ferramentas proprietárias, o que geralmente pode inviabilizar financeiramente sua utilização por outros setores ou ainda customizações, quando necessárias.

4.3 GERENCIAMENTO DE DADOS COLETADOS POR RSSF

Para gerenciar os dados de RSSF, alguns trabalhos focam na extração dos dados por meio da interação com a RSSF (MADDEN, FRANKLIN, *et al.*, 2005) (YAO e GEHRKE, 2002). Este último apresenta uma abordagem baseada em tarefas por meio de consultas declarativas. Dada uma consulta do usuário, um administrador cria um plano de execução para esta indicação. Um nó líder é necessário para consolidar os dados dos outros nós. Os nós não líderes têm a função de coletar dados periodicamente e enviar ao nó líder. No entanto, destaca-se os problemas com operações de agregação, que é dependente do nó líder que está exposta à volatilidade da camada de comunicação subjacente.

Em Madden *et al.* (2005) é apresentado o TinyDB, um processador de consultas distribuídas que é executado em cada um dos nós em uma rede de sensores, especialmente na plataforma Berkley Mote, executada sob o TinyOS. Suas principais características incluem a utilização de um camada SQL onde são

especificados os dados que devem ser extraídos. Sem necessitar escrever código em C embarcado no sensor.

Ambos, Yao; Gehrke (2002) e Madden *et al.* (2005) têm o foco em um elemento importante de uma RSSF: o nó sensor, incluindo a alteração no seu código-fonte ou atribuição de responsabilidade diferente para este elemento.

Este trabalho propõe uma abordagem genérica para extrair, transformar e carregar dados de RSSF e outras fontes em relação a um DW. Ele se concentra na camada entre RSSF e DW. É independente da arquitetura ou fabricante.

Mesmo com uma ampla variedade de aplicações com RSSFs, há poucas alternativas para o gerenciamento de dados coletados por elas, orientado à tomada de decisão. Essa deficiência pode ser suprida pelo DW, tecnologia que está crescendo rapidamente e com grande apelo para ajudar na tomada de decisões estratégicas.

4.4 CONSIDERAÇÕES SOBRE O CAPÍTULO

Após a introdução aos conceitos RSSF e DW nos capítulos anteriores, este capítulo apresentou os trabalhos que utilizam essas tecnologias no contexto da monitoração de animais e gerenciamento de dados coletados por RSSF de maneira eficiente.

O Capítulo 4 reúne resumos de pesquisas nessas áreas destacando a existência de outras soluções para o gerenciamento de dados de RSSF, porém enfatizando a ideia da utilização de DW, como uma solução robusta, extensível e independente de fabricante e sistema operacional existente nos nós das RSSFs.

5 ARQUITETURA PROPOSTA

5.1 VISAO GERAL

A arquitetura proposta nesse trabalho foca no gerenciamento dos dados de monitoração de polinizadores, coletados por RSSFs com DW. Cabe ressaltar que a utilização do DW se deve à característica do mesmo de consolidar grande volume de dados, mesmo que oriundos de diferentes fontes, com a expectativa de que uma vez integrados possam ajudar os interessados em obter respostas a perguntas complexas sobre o negócio ou ainda tomar decisões com base nos relatórios e gráficos que podem ser gerados.

Sendo assim, a Figura 12 apresenta a arquitetura proposta neste trabalho, que considera dados de RSSF e dados de outras fontes, não restringindo que tipos de fontes sejam esses, porém esperando-se que elas sejam relacionadas com a visão do negócio que se deseja analisar.

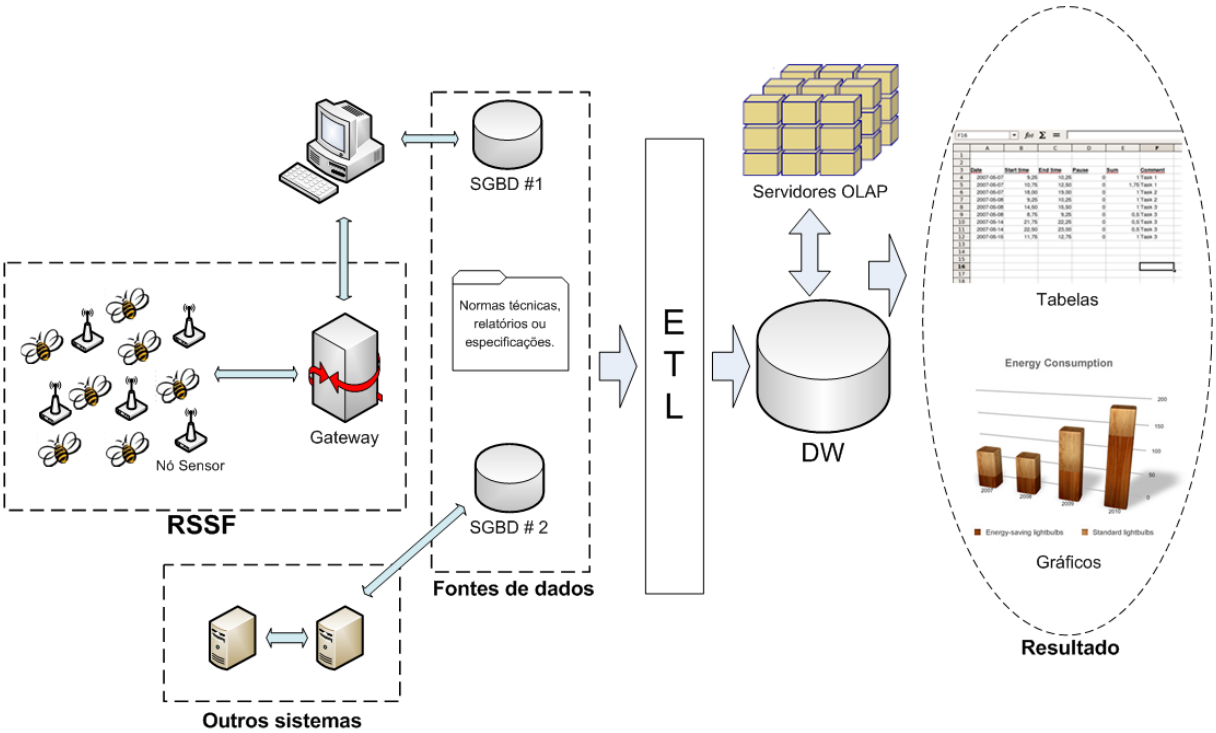


Figura 12. Arquitetura proposta.

À esquerda da Figura 12 observam-se as RSSFs que monitoram colônias de abelhas, ilustrando a coleta de dados sobre temperatura e umidade. A próxima seção detalhará os dados das RSSFs.

Outros sistemas também são representados na arquitetura, com o objetivo de permitir que dados relevantes para o contexto monitorado sejam utilizados no DW, de maneira que melhorem a confiabilidade das visões geradas pelo OLAP, permitindo, assim, dados com mais exatidão e informações mais detalhadas para os usuários finais do sistema, geralmente os tomadores de decisão. A seção 5.3 apresenta mais detalhes sobre os dados de outros sistemas.

Para completar as fontes de dados que a arquitetura apresentada suporta, tem-se uma representação para Normas Técnicas, relatórios ou especificações. Tais tipos de dados não necessariamente são oriundos do próprio experimento, e podem vir de órgãos ou instituições relacionadas ao sistema monitorado que emitam documentos que possam ser utilizados no sistema proposto, e possam ser analisados em conjunto com os dados coletados pelas demais fontes acima apresentadas. Um exemplo hipotético para esse tipo de dados é um Relatório Técnico criado por uma ONG preocupada com o aquecimento global, que monitorou e tabulou dados sobre temperatura em algumas regiões do país, uma vez a cada mês durante cinco anos e assim pôde analisar o aumento da temperatura no país. Esses dados podem ser cruzados com os dos experimentos para analisar como esse aumento global da temperatura pode ter influenciado no experimento, ou se não houve interferência.

5.2 POR QUE RSSFs?

O Capítulo 1 destaca diversas pesquisas científicas que utilizam RSSFs para monitorar animais, ressaltando as características dessa tecnologia e a utilidade da mesma em experimentos científicos. Destaca também o comportamento dos nós sensores quando espalhados em determinado *habitat*, onde são responsáveis por coletar um volume dados que precisam ser tratados para gerar informação útil para os tomares de decisão.

Contudo, atualmente os experimentos que utilizam RSSFs para monitorar ambientes ou fenômenos estão limitados aos dados coletados pelos sensores, que geralmente são apresentados na forma de planilha de texto por *softwares* que são comercializados junto com os equipamentos. Tais soluções são funcionais, porém de aplicabilidade restrita, não permitindo flexibilidade na apresentação desses dados.

A Tabela 2 apresenta um exemplo de dados coletados em um experimento com abelhas, que refletem uma imagem do ambiente monitorado no tempo apresentado. Tais dados são armazenados em linhas (tuplas) em um banco de dados relacional.

Tabela 2. Dados monitorados em experimentos com abelhas.

Nó #	**Tempo**	**Temp. interna**	**Temp. externa**	**Umidade interna**	**Umidade externa**
1	10/11/2009 7:30:10	20,3°	25,1°	70%	80%
2	10/11/2009 7:30:10	19,9°	25°	66%	83%
3	10/11/2009 7:30:20	21°	24°	65%	81%
4	10/11/2009 7:30:20	20,5°	24,7°	60%	79%

Observa-se que os dados possuem um intervalo entre as amostras coletadas, que podem variar de sensor para sensor, pois cada um deles possui programação específica. No exemplo acima o intervalo é de 10 s. Assumindo que o experimento possui quatro nós sensores e que os dados são monitorados durante 24 horas por dia, tem-se, para apenas um dia, cerca de 34.560 registros inseridos no banco de dados, o que tornaria praticamente inviável a análise de todos esses dados sem o auxílio de alguma ferramenta de tecnologia de informação. A seção 5.6 compara a solução proposta, que envolve o uso do DW, como ferramenta de tecnologia de informação, com a solução tradicional que se assume que seja a análise manual dos dados em planilhas.

5.3 OUTROS SISTEMAS (OUTRAS FONTES)

Dados de outras fontes podem auxiliar os interessados no experimento a obter informação mais precisas sobre o ambiente monitorado. No casa da monitoração de abelhas, várias outras variáveis podem ser consideradas ao analisar o comportamento desses animais.

Tais variáveis podem ser obtidas por meio de redes LonWorks, estações meteorológicas, câmeras de vídeos e outros experimentos integrados, como é o caso do contador de abelhas (CUGNASCA, 2008).

No caso dos dados oriundos de estações meteorológicas no Brasil, destaca-se o Instituto Nacional de Pesquisas Espaciais, que tem a missão de produzir ciência e tecnologia nas áreas espacial e do ambiente terrestre e oferecer produtos e serviços em benefícios do Brasil (INPE, 2011). O instituto disponibiliza por meio do seu sítio na internet, dentre outras informações, dados históricos sobre temperatura, índice de precipitação de diferentes estações meteorológicas espalhadas em todo o país.

Os dados históricos podem ser consultados e utilizados como fontes auxiliares para outros estudos. A Figura 13 apresenta a média de temperatura por mês de uma estação meteorológica localizada em São Paulo – SP. Os dados são referentes ao ano de 2010.

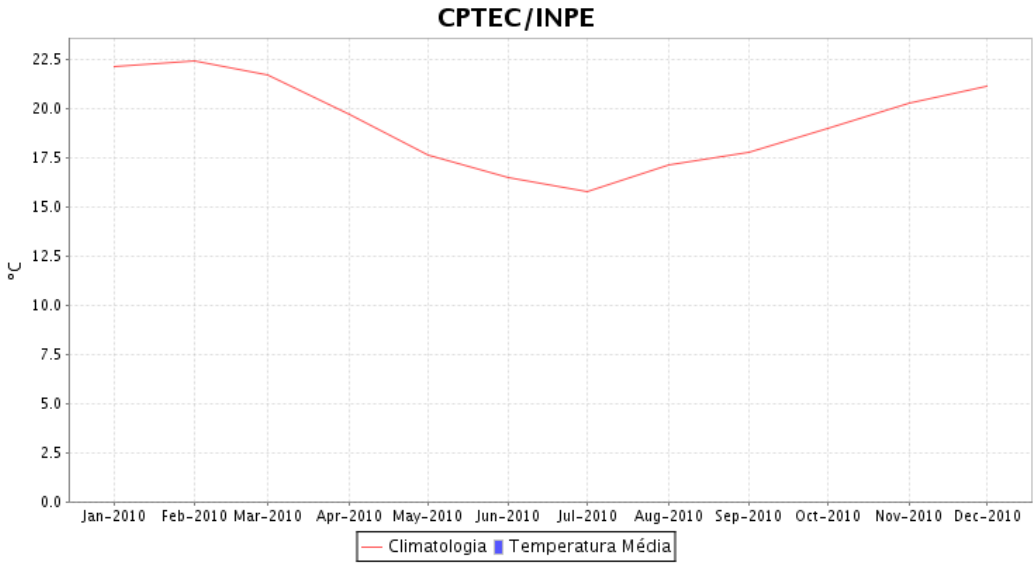


Figura 13. Dados INPE - Temperatura média 2010 - São Paulo.

5.4 INTEGRAÇÃO DOS DADOS

A etapa seguinte apresentada na arquitetura é a integração dos dados que compõem as fontes de dados do sistema proposto. Tal atividade está relacionada ao processo ETL do DW, por meio do qual os dados são extraídos de suas fontes, transformados e carregados no DW.

Os dados oriundos de base de dados devem ser obtidos por meio de comandos SQL e transformados de acordo com a granularidade deseja no modelo. O mesmo ocorre com os dados tabulados em planilhas de texto, que devem ser extraídos e normalizados. Para a extração pode-se utilizar códigos escritos em linguagem Java que leem esses arquivos e obtêm seus dados.

5.5 CONSIDERAÇÕES SOBRE O CAPÍTULO

Este capítulo apresentou a arquitetura proposta nesse trabalho, destacando a flexibilidade da mesma em suportar dados de outras fontes e a importância que eles podem apresentar no refinamento dos resultados e na melhoria na análise pelos usuários interessados no sistema.

O Capítulo 6 abordará um caso de uso, no qual foi elaborado um experimento para prova de conceito, baseada na arquitetura proposta nesse Capítulo.

6 CASO DE USO

Esta seção apresenta um caso de uso para a arquitetura proposta no capítulo anterior, validado por meio da construção de um protótipo funcional, que se baseou nos dados gerados nos experimentos com abelhas obtidos por pesquisadores de alguns grupos de pesquisa.

6.1 PROJETO VINCES

O *Virtual Network Center of Ecosystem Services* (ViNCES) é um consórcio de laboratórios de pesquisa com foco em serviços de ecossistema, tais como fotossíntese e polinização. Tem como objetivo promover um melhor entendimento sobre os serviços ambientais, sua importância, magnitude e impacto, a partir de diferentes pontos de vista - científico, econômico e social - os quais poderão conduzir a uma utilização sustentável dos recursos naturais envolvidos (VINCES, 2010).

Dentro do projeto ViNCES destaca-se o weblab[1] com foco em polinizadores que permite a integração dos conteúdos e experimentos a um conjunto de laboratórios de pesquisa por meio da Internet. O principal objetivo deste weblab é a monitoração de uma colmeia de abelha sem ferrão (*Meliponini*), com a intenção de pesquisar a influência das mudanças climáticas no comportamento das abelhas (VINCES, 2010).

O projeto se desenvolveu em parceria do Laboratório de Automação Agrícola - LAA da Escola Politécnica da Universidade de São Paulo – EPUSP, Laboratório de Abelhas do Instituto de Biociências da USP – IBUSP, Seção de Fisiologia e Bioquímica de Plantas do Instituto de Botânica, Secretaria de Estado do Meio

[1] São laboratórios nos quais experimentos, equipamentos e sistemas reais podem ser monitorados e controlados à distância por meio da Internet. Seu intuito é viabilizar o compartilhamento de equipamentos e ideias entre pesquisadores de diferentes laboratórios.

Ambiente - SFBP-IBt e Universidade da Califórnia (San Diego) dos EUA. Ele contou com o apoio da São Paulo (FAPESP), no contexto do Projeto TIDIA/KYATERA.

6.2 TERMORREGULAÇÃO

Dentre os indicadores monitorados no projeto ViNCES destacam-se as variáveis ambientais locais (temperatura e umidade), pois elas estão diretamente ligadas à termorregulação colonial.

Em Neto (2010) é apresentado uma análise sobre o mecanismo de termorregulação das abelhas sem ferrão (*Melipona quadrifasciata anthidioides*) criadas em cativeiro em Ribeirão Preto, SP. A pesquisa considera as variações de temperatura e umidade relativa do ar em diferentes locais da colmeia. Destacam-se as RSSFs e redes LonWorks para obtenção desses dados.

6.3 PROTÓTIPO FUNCIONAL

O protótipo funcional apresentado a seguir baseia-se na análise dos dados coletados pelo experimento de termorregulação do projeto ViNCES.

6.3.1 Modelagem dos dados

Os dados monitorados são temperatura e umidade, ambos coletados em diferentes locais da colmeia: Entrada, Região de Cria e Caixa de controle. E ainda, no ambiente do experimento. Na Figura 12 o bloco RSSF representa justamente esses dados coletados.

Uma amostra dos dados coletados sobre temperatura é apresentada na Tabela 3.

Tabela 3. Exemplo de dados sobre temperatura em °C. Extraído de (NETO, 2010).

Data/Hora	Entrada	Cria	Ambiente	Controle
2010-02-04 00:04:27	26,8	30,09	23,5	24,83
2010-02-04 00:09:27	26,72	30,09	23,55	24,76
2010-02-04 00:14:27	26,68	30,08	23,38	24,71
2010-02-04 00:19:27	26,66	30,05	23,34	24,66
2010-02-04 00:24:27	26,62	29,98	23,41	24,62
2010-02-04 00:34:27	26,54	29,83	23,24	24,53

A Tabela 4 apresenta uma amostra de dados sobre umidade.

Tabela 4. Exemplo de dados sobre umidade em %. Extraído de (NETO, 2010).

Data/Hora	Entrada	Cria	Ambiente	Controle
2010-02-04 00:04:26	78,4	71	85	55
2010-02-04 00:09:26	78,1	71	85	55
2010-02-04 00:14:26	78,1	71	85	55
2010-02-04 00:19:26	77,8	71	86	55
2010-02-04 00:24:26	77,9	71	85	55
2010-02-04 00:29:26	77,6	72	86	55
2010-02-04 00:34:26	77,4	72	86	55

Ambos os conjuntos de dados foram obtidos por meio de RSSFs que coletam amostras a cada 5 min. A Figura 14 apresenta fotos dos sensores utilizados, já descritos no Capítulo 3 (MEMSIC, 2010) instalados nas colmeias.

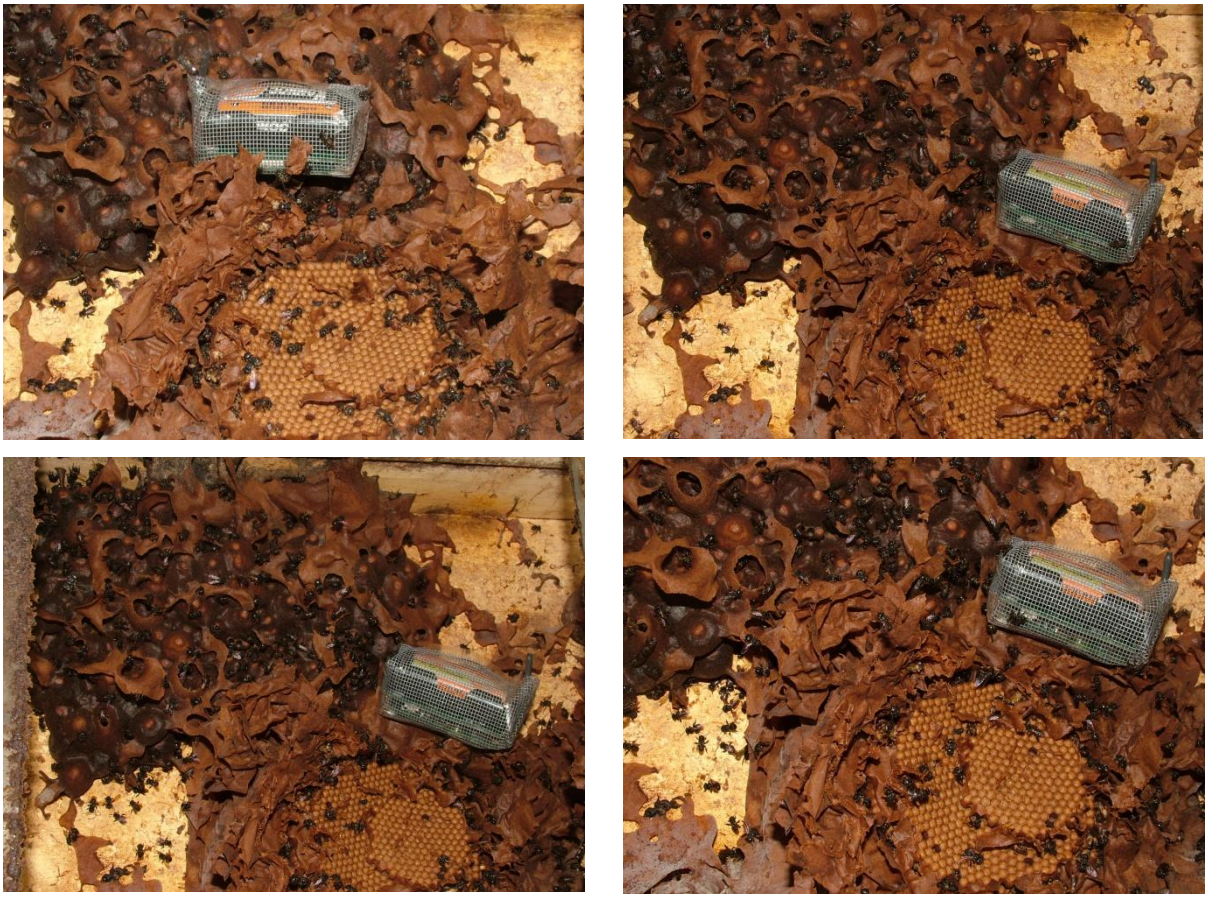


Figura 14. Fotos de nós sensores coletando dados. Extraído de (NETO, 2010).

6.3.2 Extração, Transformação e Carga dos dados

No cenário descrito os dados coletados e analisados são armazenados em banco de dados, modelado segundo a abordagem Entidade Relacionamento (ER) ou ainda, estão armazenados em planilhas eletrônicas.

Uma das contribuições deste trabalho é um processo de extração dos dados das fontes suas, seguido de uma transformação e normalização dos mesmos para que posteriormente sejam armazenados num modelo dimensional, conforme a Figura 15 ilustra.

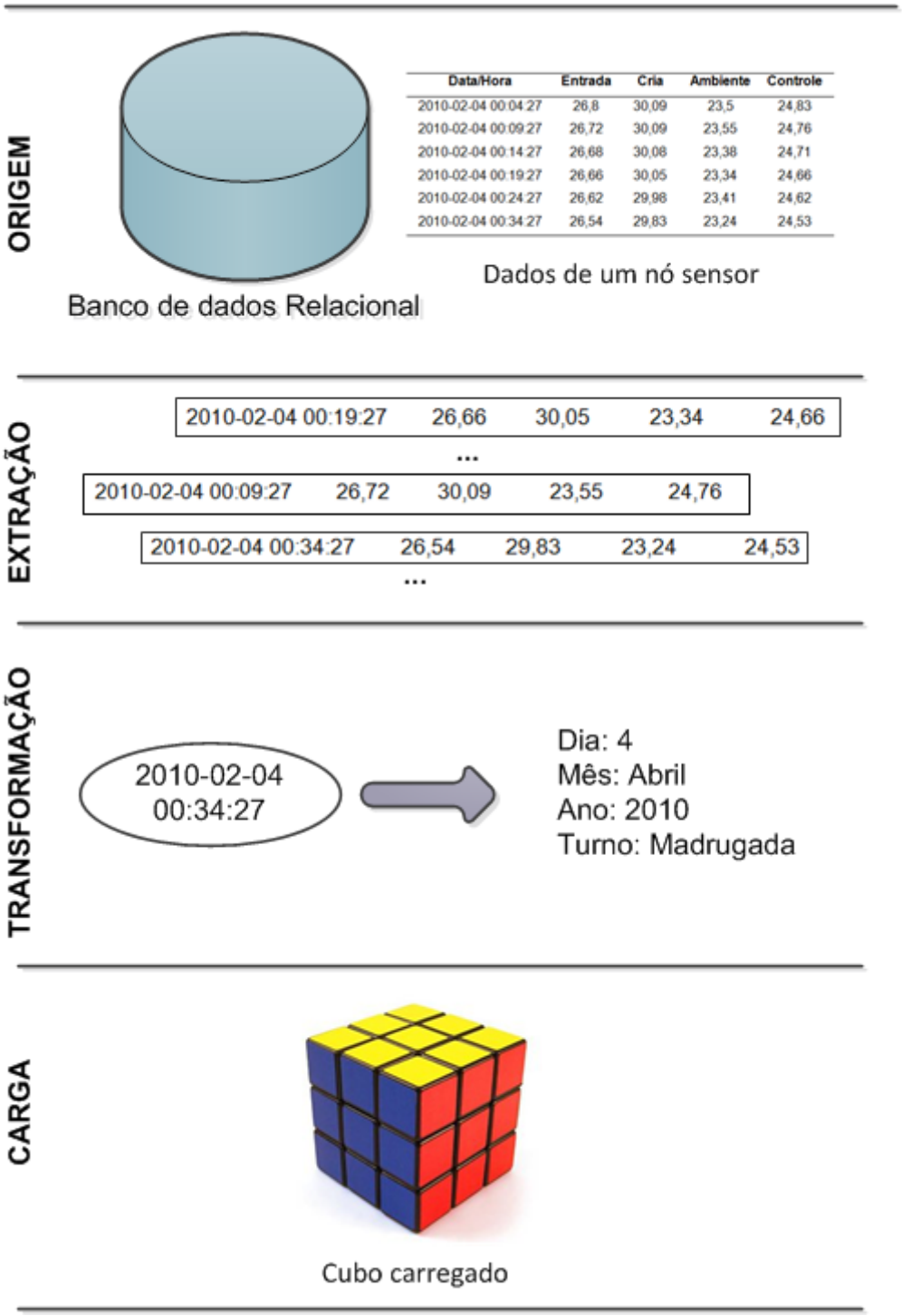


Figura 15: Extração, Transformação e Carga dos dados coletados.

O processo de extração dos dados pode variar de acordo com essa fonte dos dados. Para os dados armazenados em base de dados, a extração é feita por meio de comandos SQL, a transformação com auxílio de funções do SQL e a carga por meio de chamadas INSERT.

O Apêndice B contém o script SQL responsável pelas três etapas do modelo, extração, transformação e carga dos dados na tabela fato Temperatura.

Para a transformação dos dados sobre tempo e turno usa-se um outro script SQL auxiliar (Apêndice B) que divide em turno as amostras coletadas a cada 5 min, obtendo uma média de todos os valores do turno referente.

6.3.3 Dados de estação meteorológica

O protótipo funcional é alimentado com dados de outras fontes, conforme a arquitetura proposta prevê. Neste caso, os dados são sobre a temperatura, e são observados por uma estação meteorológica do INPE. Tais dados foram cruzados com os dados coletados pelas RSSFs a fim de melhorar a representação final dos dados, deixando-a mais apurada.

Diferente dos dados das RSSFs que necessitam de comandos SQL para serem extraídos, os dados oriundos das estações meteorológicas são fornecidos em planilhas de texto e para sua extração usa-se código Java capaz de ler esse tipo de arquivo, extrair os dados, transformá-los e em seguida carregá-los no DW.

6.3.4 Modelo Dimensional

Para criação do protótipo funcional do DW faz-se necessário a concepção de um modelo dimensional que considere as dimensões de interesse dos especialistas e observem os fatos que se deseja analisar.

A Figura 16 apresenta um modelo dimensional para o fato Temperatura, que considera as dimensões:

- **Localidade**: região onde o experimento ocorre. Por exemplo: Ribeirão Preto, Manaus e Porto Alegre.
- **Colônia**: dados sobre a colônia monitorada. Número da colmeia e uma breve descrição.

- **Turno**: madrugada, Manhã, Tarde e Noite. Essa dimensão permitirá aos interessados analisar o comportamento das abelhas em diferentes momentos do dia.
- **Espécie**: indica a espécie da abelha monitorada.
- **Tempo**: indica o ano, mês e dia.

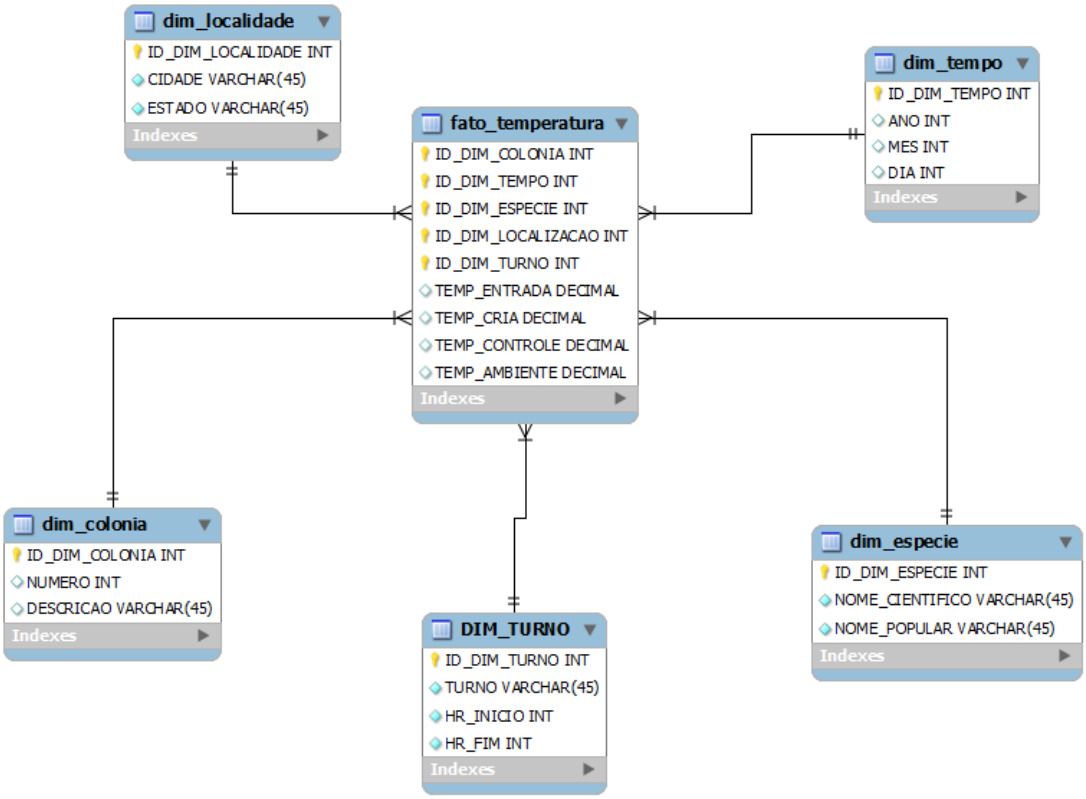


Figura 16. Modelo dimensional proposto.

A matriz de barramento para o modelo proposto é representada na Figura 17.

Processos de Negócios (Tabelas Fato)	**Dimensões**				
	Localidade	Colônia	Turno	Espécie	Tempo
Temperatura	x	x	X	x	x
Umidade	x	X	X	x	x

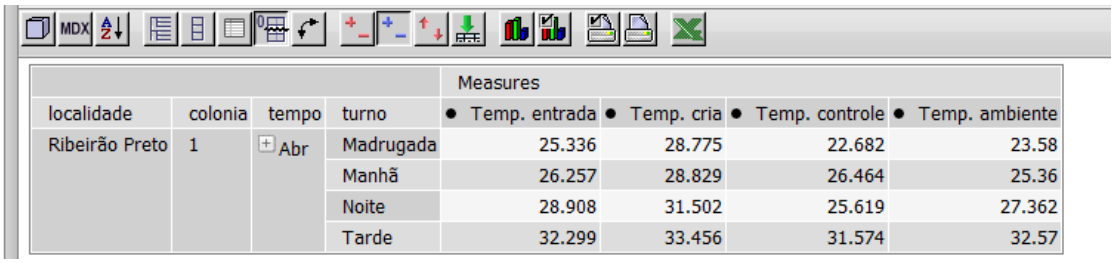
Figura 17. Matriz de barramento do modelo proposto.

6.3.5 Aplicação da Ferramenta OLAP

A ferramenta OLAP utilizada para criação do DW é baseada no na linguagem de programação de *scripting* Mondrian, que foi desenvolvida pela Universidade de Utrecht para ser utilizada em aplicações da Internet. Ele se constitui em um servidor OLAP, sendo um dos principais componentes da plataforma Pentaho de *Business Intelligence*. Ele atua como um conector entre uma interface que interage diretamente com o usuário, baseado em Java, e um banco de dados relacional, uma espécie de conector JDBC para OLAP, gerando *queries* SQL para o banco de dados e processando os dados resultantes. A sua escolha se deve ao seu alto desempenho, possibilitando a análise interativa de volumes grandes ou pequenos de informação, e também por ser de código aberto (MONDRIAN, 2010). O Apêndice A apresenta passo a passo para instalação e configuração desse ambiente.

A facilidade de navegação e de cruzamentos proporcionada pelo DW permite a obtenção dinâmica dos dados, possibilitando ainda a inclusão e exclusão de variáveis de forma rápida pelo próprio pesquisador/especialista/interessado na solução; a ferramenta ainda oferece recursos de navegação interna sobre os registros, exportação para outros formatos como “xls” do Excel e “pdf”, e geração dinâmica de gráficos (apresentados com mais detalhes no Apêndice A).

A Figura 18 apresenta uma das telas do protótipo construído do DW. Tal visão serve de passo inicial para os interessados nas análises customizarem a pesquisa e emitirem relatórios.



				Measures			
localidade	colonia	tempo	turno	Temp. entrada	Temp. cria	Temp. controle	Temp. ambiente
Ribeirão Preto	1	Abr	Madrugada	25.336	28.775	22.682	23.58
			Manhã	26.257	28.829	26.464	25.36
			Noite	28.908	31.502	25.619	27.362
			Tarde	32.299	33.456	31.574	32.57

Figura 18. Tela do protótipo criado utilizando o Mondrian/Pentaho.

Além disso, é possível obter gráficos baseados nos filtros definidos e exportar em formato planilha de texto (Figura 19) e gráfico (Figura 20).

	A	B	C	D	E	F	G	H
1					Measures			
2	localidade	colonia	tempo	turno	Temp. entrada	Temp. cria	Temp. controle	Temp. ambiente
3	Ribeirão Preto	1	Abr	Madrugada	25,336	28,775	22,682	23,58
4				Manhã	26,257	28,829	26,464	25,36
5				Noite	28,908	31,503	25,619	27,363
6				Tarde	32,299	33,456	31,574	32,57
7								

Figura 19. Planilha eletrônica gerada pelo Mondrian/Pentaho.

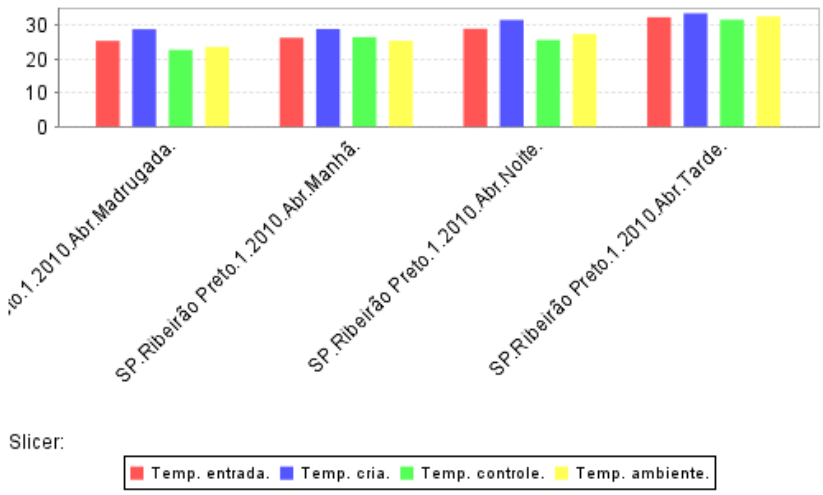


Figura 20. Gráfico gerado pelo Mondrian/Pentaho.

E ainda, com auxílio da ferramenta OLAP é possível obter relatórios flexíveis que respondam perguntas específicas do experimento, como por exemplo: Qual a média de temperatura na entrada da colmeia nas noites no período de 16 a 19 de Abril de 2010? A Figura 21 responde a essa pergunta.

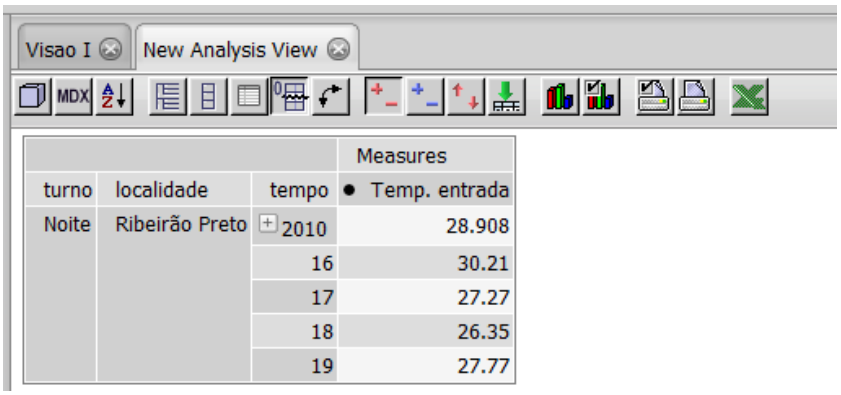


Figura 21: Tela do protótipo com visão customizada.

6.4 COMPARATIVO DAS TÉCNICAS

Esta seção compara a solução proposta, que utiliza o DW, com a solução tradicional, baseada na consulta em planilhas de texto ou base de dados.

A extração de informações a partir de banco de dados ou planilha eletrônica é uma atividade que pode custar bastante tempo, variando de acordo com os detalhes desejados. E ainda, para que se obtenha informação realmente interessante ao negócio e que dê suporte à tomada de decisões por parte dos interessados, os dados ainda precisam ser processados.

Partindo do princípio que os dados não são processados, no caso da planilha eletrônica, o usuário (supondo que ele possua um conhecimento mediano em *softwares* de planilha eletrônico) pode gera gráficos, obter médias, variâncias, máximas e mínimas, aproveitando os recursos que o editor de planilhas oferece. Esse processo de tratamento dos dados pode custar um tempo significativo, dependendo do conhecimento do usuário e, além disso, tais ações podem ser aplicáveis somente para um conjunto de dados em questão, ou seja, não sendo permitida a comparação com dados históricos, por exemplo.

Já quanto aos dados armazenados em um banco de dados, estes precisam ser tratados por uma aplicação, que por meio de comandos SQL realize consultas e obtenha informações pertinentes e úteis para os pesquisadores e demais interessados. Porém, é nesse ponto que o DW (modelagem dimensional) se destaca do banco de dados tradicional (modelo relacional), pois a maneira como o modelo de dados foi concebido e consequentemente, a maneira como os dados estão armazenados no banco, permite que o DW implementado extraia informações com mais exatidão e ainda atendam aos requisitos do sistema desejado.

Cabe destacar que para o sucesso do DW é muito importante o envolvimento de todos os interessados, no processo de concepção do modelo de dados, etapa de análise de requisitos. Isso diminui a possibilidade de que ao final da definição do modelo do DW, execução do processo de ETL e extração de dados e relatórios, perguntas importantes sobre o negócio não sejam respondidas. Há casos que isso pode acontecer, devido a uma falha na etapa de análise dos requisitos. Além disso, um bom DW pode permanecer em funcionamento durante vários anos, sem qualquer manutenção, basta que a rotina de ETL seja configurada para isso.

Após essa análise pode-se sumarizar que a maneira como os dados oriundos das RSSFs serão tratados para gerar informação aos interessados, pode variar bastante, sendo que as principais são as planilhas eletrônicas e os bancos de dados. Esse trabalho apresenta uma prova de conceito com um DW atuando com agregador dos dados das RSSFs e demais dados de interesse do experimento. E pode-se perceber que, se o modelo dimensional do DW for bem desenvolvido, ouvindo-se todas as partes interessadas, pode-se chegar a uma solução que atenda a expectativa dos usuários e, gere aumento na produtividade, pois os dados que antes necessitavam ser processados em planilhas ou programas para obter informações de negócio são periodicamente carregados no DW, por meio do processo de ETL. Além disso, é possível ainda obter visões customizadas sobre o negócio, respondendo perguntas importantes e ainda possibilitando comparações com dados históricos.

6.5 CONSIDERAÇÕES SOBRE O CAPÍTULO

Este capítulo apresentou as principais características do protótipo funcional criado para validar a arquitetura apresentada no Capítulo 5. A validação por meio do protótipo permitiu verificar a solução como ferramenta para geração de relatórios flexíveis e ainda destacou o os benefícios da utilização do DW em comparação às planilhas de textos de banco de dados relacionais. Cabe destacar que o sistema proposto está preparado para receber dados de experimentos de outras localidades e de diferentes espécies de abelhas.

7 CONSIDERAÇÕES FINAIS

A prova de conceito criada extraiu informações de experimentos realizados na cidade de Ribeirão Preto em São Paulo, apresentando para o pesquisador e demais interessados uma ferramenta de fácil navegação e extração de dados. Porém, o presente trabalho ressalta a flexibilidade da solução em gerenciar dados de outras cidades e experimentos.

O modelo dimensional criado para o estudo de caso apresenta as dimensões: Tempo, Localidade, Turno, Colônia e Espécie e indicadores, que são variáveis de cruzamento de informações relevantes para os interessados nos experimentos.

A criação do DW permitiu aos pesquisadores ter uma agregação de dados de outras fontes, como é o caso dos dados da RSSF e estações meteorológicas. Conforme destacado, esse processo de agregação de dados é muito importante para melhoria da exatidão dos dados, porém é uma atividade que pode consumir algum tempo dentro do processo de ETL, pois é necessário o nivelamento dos dados que podem estar em diferentes granularidades.

Outro fato importante é a utilização da dimensão Turno, que permite aos usuários analisar o comportamento das abelhas em diferentes horários do dia e ainda analisar esses intervalos de horas considerando uma variação de meses, ano ou estações climáticas.

Entretanto, destaca-se que esses ainda são passos iniciais para que seja permitido afirmar que as informações disponíveis para a criação de abelhas estão próximas das formas de obtenção e acesso de informações existentes em outras áreas, como exemplo a área financeira.

Em resumo, o trabalho apresentou a utilização DW para gerenciar dados coletados por RSSF e outras fontes que monitoram abelhas. Revisão da literatura sobre os diversos temas relacionados e caso de uso com criação de protótipo foram apresentadas e discutidas.

7.2 CONTRIBUIÇÕES

Este trabalho apresenta algumas contribuições para o contexto de pesquisa do Laboratório de Automação Agrícola da EPUSP e do IBUSP, relacionando-se também com outros trabalhos e projetos.

Dentre as principais contribuições, podem-se citar:

- **Ferramenta de auxílio eficaz à avaliação de dados de experimentos:** o modelo proposto apresenta um bom potencial para possibilitar a análise metódica e simplificada de experimentos que geram grande volume de dados. O advento das RSSFs vem proporcionando novas experiências na coleta de dados, possibilitando a obtenção rápida e simplificada de muitos dados. A sua evolução certamente colocará um novo desafio aos pesquisadores, que é a busca por formas eficazes de análise de grande quantidade de dados. Nesse sentido, a ferramenta proposta e discutida nesse trabalho mostra grande potencial de contribuir na superação desses desafios.

- **Uso de parâmetros de experimentos realizados em laboratório em Ribeirão Preto:** esta pesquisa utilizou dados de experimentos com RSSF que monitoram abelhas em laboratórios da USP em Ribeirão Preto, após o tratamento dos dados e carga no DW, relatórios e gráficos foram gerados sobre a monitoração de abelhas, o que gera retorno rápido aos pesquisadores, que podem considerar tais informações no momento de reprogramar seu experimento.

- **Flexibilidade da solução:** a solução apresentada mostrou-se flexível para agregar dados de outros experimentos, seja em outras localidades, como, por exemplo, em Manaus, onde há experimentos com abelhas sem ferrão, porém de espécie diferente. E ainda, o sistema proposto pode ser adaptado para outros cenários, como a agricultura em geral, como o cultivo de milho, uva, entre outros, e em especial a agricultura de precisão. A facilidade para a adaptação para as novas necessidades e até mesmo para outros contextos constitui-se em um dos méritos da proposta apresentada.

- **Rapidez na tomada de decisões:** outro ponto que merece destaque é a maior rapidez na tomada de decisões. Em muitas pesquisas, como as realizadas com abelhas e agricultura, devem ser realizadas em épocas específicas. A maior rapidez na análise dos dados e a partir disso alterar a condução dos novos experimentos é um fator importante para aumentar o sucesso das pesquisas. Isso pôde ser comprovado nos experimentos com RSSFx que monitoram abelhas em laboratórios da USP em Ribeirão Preto, nos quais após o tratamento dos dados e carga no DW, relatórios e gráficos foram gerados sobre a monitoração de abelhas, gerando um retorno rápido aos pesquisadores, que puderam considerar tais informações no momento de reprogramar seus experimentos. Assim, a alternativa do DW como ferramenta para auxílio na rápida tomada de decisões é um ponto de destaque da solução proposta.

7.3 TRABALHOS FUTUROS E MELHORIAS

Muitas questões importantes foram apenas citadas, tratadas de maneira introdutória ou nem mesmo mencionadas. Novos trabalhos de pesquisa, como continuidade deste, poderiam abordá-las mais profundamente. Algumas delas são:

- Expansão do projeto para outros experimentos que utilizam RSSF ou não. Dentro do próprio Projeto ViNCES existiram outro projetos que poderiam ter usufruído dos benefícios de um DW para auxiliar os pesquisadores em tomar rápidas decisões, como é o caso do weblab sobre fotossíntese. Além disso, o LAA também possui várias pesquisas com dados tabulados coletados por RSSF, dentre eles destacam-se aqueles relacionados à viticultura e cultura de milho.
- Busca-se ainda a carga de mais dados no sistema proposto, de maneira que o mesmo tenha uma base histórica de vários anos, o que permitirá aos interessados extrair informações ainda mais importantes e comparáveis ao longo dos anos.

- E ainda, generalizar o protótipo proposto nesse trabalho para lidar com qualquer experimento com RSSF e assim ampliando a aplicabilidade do trabalho.

Este trabalho apresentou a utilização do DW como alternativa para o gerenciamento dos dados coletados por RSSF, optando por apenas elencá-lo como mais uma opção dentre as já tradicionais soluções e assim deixando ao leitor definir a que melhor atende ao seu cenário. Porém, destaca o ganho nas tomadas de decisões estratégicas como um dos maiores motivadores dessa abordagem que ainda permite flexibilidade na extração de relatórios, respostas rápidas para questões de interesse do negócio e ainda comparação com dados históricos dos experimentos monitorados.

8 REFERÊNCIAS

AKYILDIZ, I. F. et al. Wireless sensor networks: a survey. **Computer Networks: The International Journal of Computer and Telecommunications Networking**, New York, v. 38, n. 4, p. 393-422, 2002. ISSN: 1389-1286.

BAGGIO, A. Wireless sensor networks in precision agriculture. In: WORKSHOP ON REAL-WORLD WIRELESS SENSOR NETWORKS, 2005, Stockholm. **Proceedings…** Stockholm: RealWSN, jun. 2005.

BARONTI, P. et al. Wireless sensor networks: A survey on the state of the art and the 802.15.4 and ZigBee standards. **Computer Communications: The International Journal for the Computer and Telecommunications Industry**, v. 30, n. 7, p. 1655-1695, mai. 2007. ISSN: 0140-3664.

CAMILLI, Alberto. **Um estudo sobre os aspectos da utilização de redes de sensores sem fio em aplicações para agricultura de precisão**. 2004. 180 p. Tese (Doutorado) - Escola Politécnica, Universidade de São Paulo, São Paulo, 2004.

CAMILLI, A. et al. From wireless sensors to field mapping: Anatomy of an application for precision agriculture. **Computers and Electronics in Agriculture**, Amsterdam, Holanda, v. 58, n. 1, p. 25-36, ago. 2007. ISSN: 0168-1699.

CHAUDHURI, S.; DAYAL, U. An overview of data warehousing and OLAP technology. **ACM SIGMOD Record**, New York, v. 26, n. 1, p. 65-74, mar. 1997.

CROSSBOW. **Crossbow Technology**, jan. 2010. Disponível em: <www.xbow.com>. Acesso em: 1 ago. 2010.

CUGNASCA, C. E. Redes de controle e redes de sensores sem fio aplicadas às pesquisas com meliponíneos. In: ENCONTRO SOBRE ABELHAS, 8., Ribeirão Preto, São Paulo. **Anais...** Ribeirão Preto: FUNPEC, 2008. p. 11-11.

ELMASRI; N. **Sistema de Banco de Dados**: Fundamentos e Aplicações. 4. São Paulo: Pearson Education, 2005.

ESTRIN, D. et al. Instrumenting the world with wireless sensor networks. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL

PROCESSING, 2001, Salt Lake City. **Proceedings…** Salt Lake City: IEEE, 2001, p. 2033-2036. ISSN 7167751.

ESTRIN, D.; GOVINDAN, R.; HEIDEMANN, J. Embedding the Internet: Introduction. **Communications of the ACM**, New York, v. 43, n. 5, p. 39-41, mai. 2000.

GANESAN, D. et al. Networking issues in wireless sensor networks. **Journal of Parallel and Distributed Computing**, v. 64, p. 799-814, 2003. ISSN: 0743-7315.

GONDA, L.; CUGNASCA, C.; CAMILLI, A. Aplicações Agrícolas para Redes de Sensores Sem Fio. In: CONGRESSO BRASILEIRO DE AGROINFORMÁTICA, 5., 2005, Londrina - PR. **Anais...** Londrina - PR: SBI-AGRO, 2005.

HEINZELMAN, W. R.; CHANDRAKASAN, A.; BALAKRISHNAN, H. Energy-efficient communication protocol for wireless microsensor networks. In: Annual Hawaii International Conference, 33., 2002, Hawaii. **Proceedings…** Hawaii: IEEE, jan. 2002.

INMON, W. H. **Building Data Warehouses**. 2. EUA: Wiley Books, 1997.

INPE. **Centro de Previsão do Tempo e Estudos Climáticos**. Instituto Nacional de Pesquisas Espaciais, 2011. Disponível em: <http://bancodedados.cptec.inpe.br/>. Acesso em: 10 dez. 2010.

INTANAGONWIWAT, C.; GOVINDAN, R.; ESTRIN, D. Directed diffusion: a scalable and robust communication paradigm for sensor Source. In: ANNUAL INTERNATIONAL CONFERENCE ON MOBILE COMPUTING AND NETWORKING, 6., 2000, Bouston. **Proceedings…** Bouston: ACM, ago. 2000. p. 56-67.

JUANG, P. et al. Energy-efficient computing for wildlife tracking: design tradeoffs and early experiences with zebranet. In: ACM SIGOPS OPERATING SYSTEMS REVIEW, 2002, New York. **Proceedings…** New York: ACM, 2002. p. 96-107.

KIMBALL, R. A Dimensional Modeling Manifesto. **DBMS - Special issue on data warehousing**. San Francisco, v. 10, n. 9, ago. 1997. San Francisco: ACM. 1997. p. 58-70.

KIMBALL, R.; ROSS, M. **The data warehouse toolkit**: the complete guide to dimensional modeling. 2. EUA: Wiley, 2002.

KINNEY, P. ZigBee Technology: Wireless Control that Simply Works. **Communications Design Conference**, 2003. Apresentado a Chair of IEEE 802.15.4 Task Group, out. 2003.

LAUDON, K.; LAUDON, J. **Sistemas de Informações Gerenciais**. 7. São Paulo: Makron Books, 2007.

LOUREIRO, A. A. et al. Redes de Sensores Sem Fio. In: SIMPÓSIO BRASILEIRO DE REDES DE COMPUTADORES, 21., 2003, Natal, **Anais...** Natal: SBC, 2003.

MADDEN, S. et al. TinyDB: an acquisitional query processing system for sensor networks. **ACM Transactions on Database Systems**, New York, EUA, v. 30, p. 122-173, 2005.

MAINWARING, A. et al. Wireless sensor networks for habitat monitoring. In: ACM INTERNATIONAL WORKSHOP ON WIRELESS SENSOR NETWORKS AND APPLICATIONS, 1., Atlanta, **Proceedings…** Atlanta: ACM. 2002. p. 97-99.

MARQUES, V. **Analisando os dados do programa de melhoramento genético**. 2002. 143 p. Dissertação (Mestrado) - Escola Politécnica, Universidade de São Paulo, São Paulo, 2002.

MEGUERDICHIAN, S. et al. Coverage problems in wireless ad hoc sensor networks. In: IEEE INFOCOM - Annual Joint Conference of the IEEE Computer and Communications Societies. 2001, Anchorage, Alaska, **Proceedings…** Anchorage, Alaska: IEEE. 2001. p. 1380-1387.

MELO, G. Comportamento social em vespas da família Sphecidae (Hymenoptera, Apoidea). **Ecologia e Comportamento de Insetos**: Série Oecologia Brasiliensis, Rio de Janeiro, v. 8, 2000, p. 85-130.

MEMSIC. **MEMSIC**: Powerful Sensing Solutions for a Better Life, 2010. Disponível em: <http://www.memsic.com/>. Acesso em: 5 ago. 2010.

MILET-PINHEIRO, P.; SCHLINDWEIN, C. A dependência mútua entre Cordia leucocephala (Cordiaceae) e Ceblurgus longipalpis (Rophitinae: Halictidae). In: ENCONTRO SOBRE ABELHAS, 8., 2008, Ribeirão Preto, **Anais…** Ribeirão Preto: USP, jul. 2008. p. 466.

MIN, R. et al. Low-Power Wireless Sensor Networks. In: INTERNATIONAL CONFERENCE ON VLSI DESIGN, 14. 2001, Bangalore, India, **Proceedings…** Bangalore, India: IEEE, jan. 2001. p. 205-210.

MINAMI, M.; MORITO, T. M. H. A Battery-less Wireless Sensor Network System for Environmental Monitoring Applications. In: INTERNATIONAL WORKSHOP ON NETWORKED SENSING SYSTEMS, 2., San Diego, CA, **Proceedings…** San Diego, CA: IEEE. 2005.

MINI, R. A.; NATH, B.; LOUREIRO, A. F. A Probabilistic Approach to Predict the Energy Consumption in Wireless Sensor Networks. In: WORKSHOP DE COMUNICAÇÃO SEM FIO E COMPUTAÇÃO MÓVEL, 4., São Paulo, 2002. **Anais...** São Paulo: ACM. 2002. p. 23-25.

MONDRIAN. **Pentaho**: Open Source Business Intelligence, Feb 2010. Disponível em: <http://mondrian.pentaho.org/>.

NAKAMURA, E.; LOUREIRO, A. Information Fusion in Wireless Sensor Networks. In: CONGRESSO DA SBC, 28., Belém, 2008, **Anais…** Belém: SBC. 2008. p. 41-48.

NETO, A. V. **Termorregulação em Melipona quadrifasciata anthidioides (Apidae: Meliponini) sob condições de alta temperature**, 2010. Dissertação (Mestrado) - Escola Politécnica, Universidade de São Paulo, Ribeirão Preto, São Paulo, 2010.

POLASTRE, J. R. **Design and Implementation of Wireless Sensor Networks for Habitat Monitoring**, 2003. Dissertação (Mestrado) - University. Berkley, USA. 2003.

RAMANATHAN, N. et al. Designing Wireless Sensor Networks as a Shared Resource for Sustainable Development. In: INTERNATIONAL CONFERENCE ON INFORMATION AND COMMUNICATION TECHNOLOGIES AND DEVELOPMENT, 1. Berkeley, EUA, 2006, **Proceedings…** Berkeley, EUA: IEEE. 2006.

SHAH, R. C. et al. Data MULEs: Modeling a Three-tier Architecture for Sparse Sensor Networks. In: IEEE SNPA WORKSHOP, Seattle, 2003, **Proceedings…** Seattle: IEEE, 2003, p. 30-41.

SILVA, F. et al. Tecnologia de nós sensores sem fio. **Revista Controle e Instrumentação**, v. 92, mai. 2004. São Paulo: Valete Editora, mai. 2004.

SOUZA, D. L.; EVANGELISTA-RODRIGUES, A.; PINTO, M. P. As Abelhas Como Agentes Polinizadores (The Bees Agents Pollinizer's). **Revista electrónica de Veterinaria**, v. 8, n. 3, mar. 2007. ISSN 1695-7504.

TINYOS. **TinyOS**, 2010. Disponível em: <http://www.tinyos.net>. Acesso em: 15 mai. 2010.

TONI, J. **Definição de um Data Mart em Cooperativas Agropecuárias**. Dissertação (Mestrado). 2002. p. 172. Universidade Federal de Florianópolis, 2002.

VINCES. **ViNCES**: Virtual Networking Center of Ecosystem Services. Site do projeto ViNCES, 2010. Disponível em: <http://www.ib.usp.br/vinces/>. Acesso em: 5 mai. 2010.

WARK, T. et al. Transforming Agriculture through Pervasive Wireless Sensor Networks. **IEEE Pervasive Computing**, v. 6, n. 2, p. 50-57, 2007.

WERNER-ALLEN, G. et al. Monitoring Volcanic Eruptions with a Wireless Sensor Network. In: EUROPEAN WORKSHOP ON WIRELESS SENSOR NETWORKS, 2., Istanbul, Turkey, 2005, **Proceedings…** Istanbul, Turkey: IEEE, 2005.

WERNER-ALLEN, G. et al. Deploying a Wireless Sensor Network on an Active Volcano. **IEEE Internet Computing**, v. 10, p. 18-25, mar. 2006. ISSN 1089-7801.

XMESH. **XMesh**: User’s Manual, 2009. Disponível em: <http://www.xbow.com/Support/Support_pdf_files/XMesh_Users_Manual.pdf>. Acesso em: 9 mai. 2009.

YAO, Y.; GEHRKE, J. The cougar approach to in-network query processing in sensor networks. **ACM SIGMOD Record**, New York, v. 31, 2002.

APÊNDICE A – CONFIGURAÇÃO MONDRIAN/PENTAHO

Este Apêndice tem como objetivo apresentar detalhes técnicos sobre a instalação e configuração do Mondrian/Pentaho (MONDRIAN, 2010), servidor OLAP de código aberto (escrito em Java) utilizado no presente trabalho.

A.1 PRÉ-REQUISITO

O ambiente pode ser configurado em sistemas operacionais Windows, Linux ou Mac OS, pois os aplicativos são todos escritos em Java, sendo assim, multiplataformas. Além disso, é necessário que o computador possua o compilador Java JDK (http://www.sun.com) na versão igual ou superior a 1.5. Feito isso, deve-se configurar com as variáveis de ambiente: JAVA_HOME e JRE_HOME definidas no CLASSPATH do sistema operacional.

A.2 INSTALAÇÃO

Após a configuração do sistema operacional, deve-se acessar o site Mondrian/Pentaho em http://mondrian.pentaho.org/ para obter os pacotes e instalação dos aplicativos:

- Pentaho BI Platform (biserver-ce-3.6.0-stable)
- Analysis View (Mondrian) (mondrian-3.2.0.13583)
- Schema – Workbench (psw-ce-3.1.6.13364)

Eles representam o servidor de aplicação Apache, o servidor OLAP Mondrian/Pentaho e o aplicativo utilizado para criação do *Schema* utilizado no Cubo.

Para a instalação deles no ambiente Windows, basta descompactar os arquivos zip em um diretório escolhido pelo usuário, e executar o arquivo start-pentaho.bat (Windows) na pasta "biserver-ce/". Para iniciar o servidor TOMCAT com o Mondrian/Pentaho, deve-se acessá-lo na URL http://localhost:8080/pentaho/Login. O usuário padrão é *Joe* e a senha é *password*. A Figura 22 apresenta a tela de entrada do Mondrian/Pentaho.

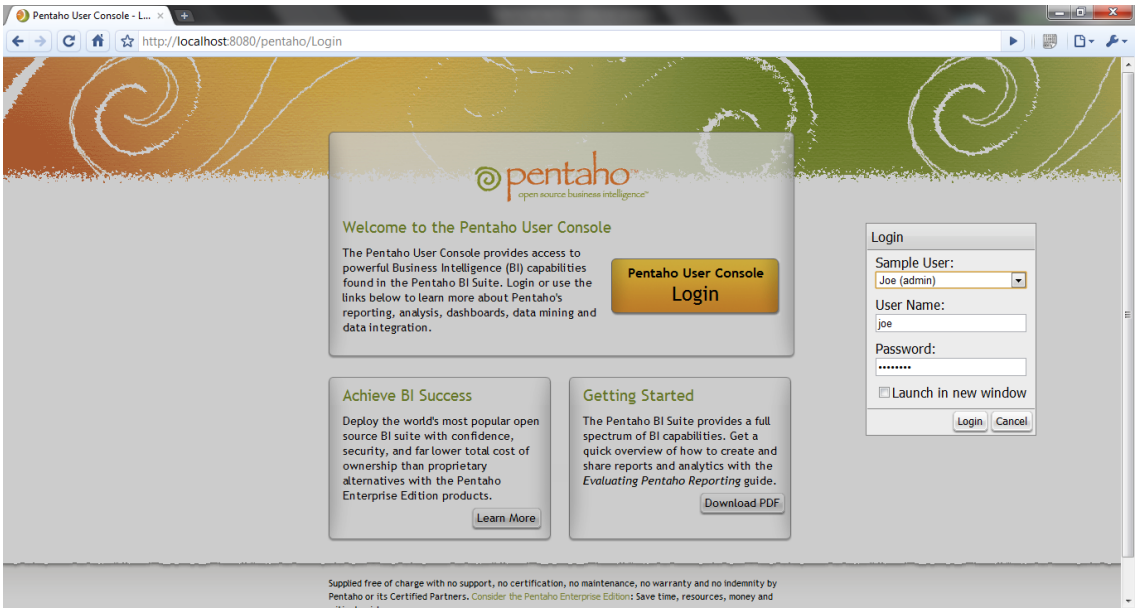


Figura 22. Tela de entrada Mondrian/Pentaho.

Após efetuar login o usuário é redirecionado para a tela inicial que contém as visões criadas pelo usuário. A Figura 23 apresenta a tela inicial da aplicação.

Figura 23. Tela inicial do Mondrian/Pentaho.

Para administração do servidor deve-se executar o arquivo start-pac.bat na pasta “administration-console/” e acessar a URL http://localhost:8099/. A Figura 24 apresenta a tela inicial do Painel de Administração do Mondrian/Pentaho.

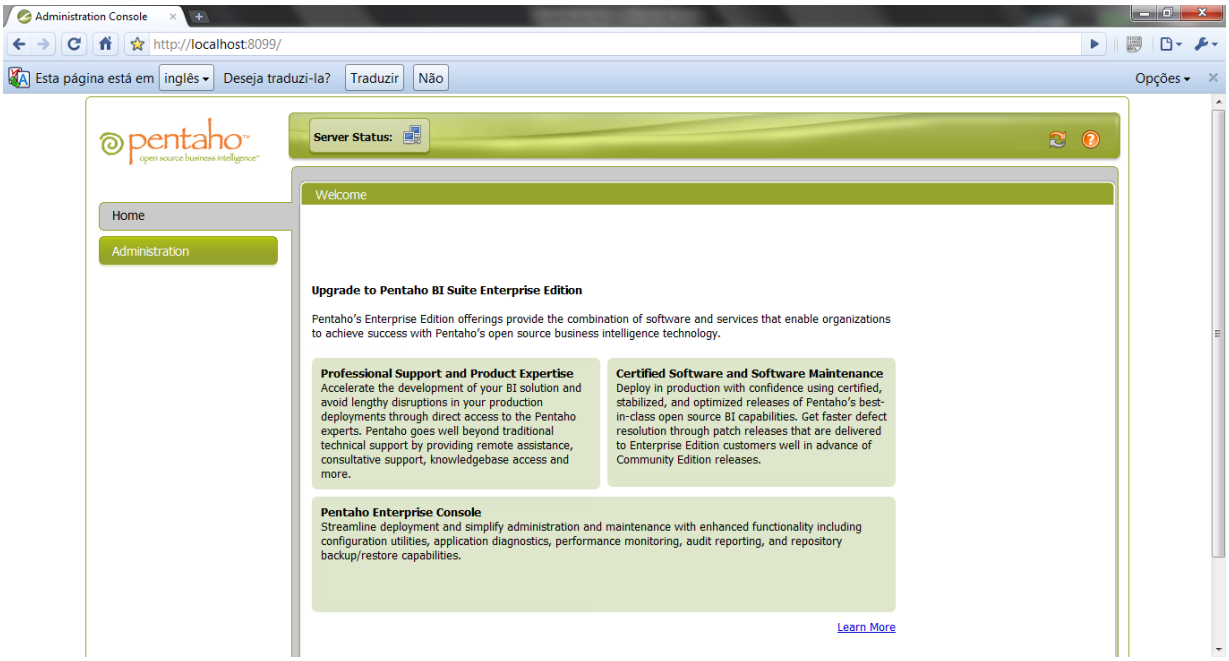


Figura 24. Tela inicial do Painel de Administração do Mondrian/Pentaho.

A.3 CONFIGURAÇÃO DO BANCO DE DADOS

O banco de dados utilizado foi o MySQL, ferramenta de código aberto disponível em http://mysql.com/.

Para utilizar o Mondrian/Pentaho com o banco de dados MySQL é necessário configurar a conexão JNDI que faz a interface entre o Java e o banco de dados, para isso é necessário o driver JDBC do MySql. O mesmo é obtido acessando a URL http://dev.mysql.com/downloads/connector/j/.

O arquivo mysql-connector-java-5.0.7.jar deve ser copiado para a pasta administration-console\jdbc e se efetuar as configurações abaixo:

Name: mysql

URL: jdbc:mysql://localhost:3306

User Name: xxx / Password: xxx

A.4 CRIAÇÃO DO SCHEMA PARA CUBO OLAP

Na etapa de criação do schema do Cubo OLAP assume-se que as tabelas já estejam criadas no banco de dados.

Para abrir a aplicação deve-se acessar o arquivo workbench.bat em schema-workbench/. Feito isso e antes de criar o schema é necessário configurar a conexão que será utilizada. A Figura 25 apresenta a tela de cadastro da conexão.

Os passos são:

1. Ir para Tools -> Connection.
2. Preencher os campos
3. Driver: org.postgresql.Driver.
4. Connection URL: jdbc:mysql://localhost:3306.
5. User name e password: root.
6. Clicar em Test Connection.
7. Clicar em Accept.

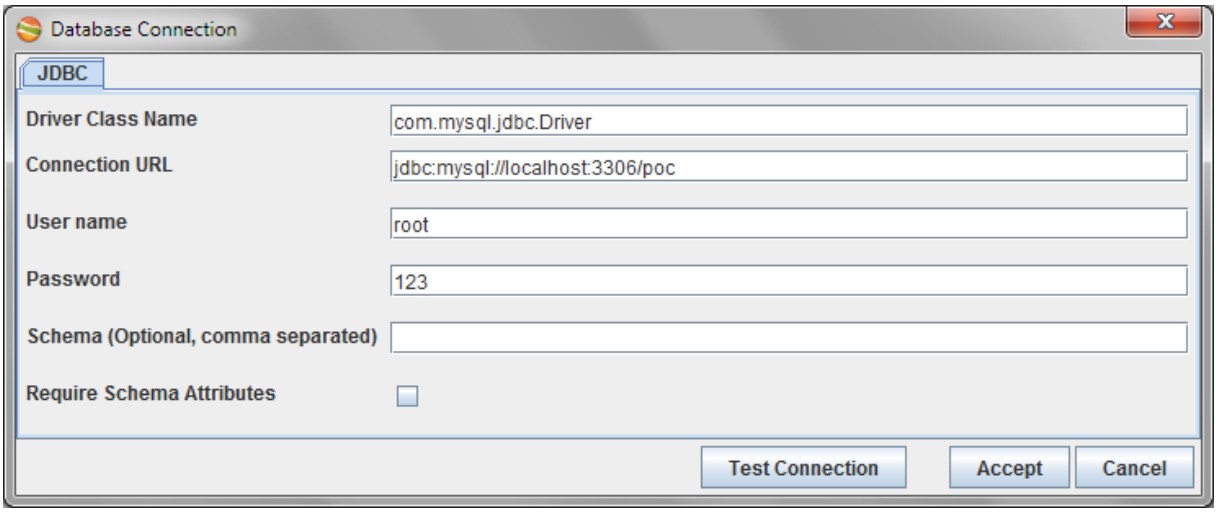


Figura 25. Configuração da conexão com banco de dados.

Após configurar a conexão com o banco de dados pode-se criar o schema com base no modelo lógico já criado no banco de dados.

1. Ir em File ->New ->Shema

2. Clicar no botão para inserir o cubo e adicione um nome que referenciará o cubo.
3. Clicar com o botão esquerdo sobre o cubo e ir em Add Table. Preenche os campos corretamente com os dados da tabela fato definida no projeto do modelo multidimensional.
4. Clicar com o botão direito do mouse sobre o cubo e ir em Add Measure. Essa entidade se referencia à métrica definida no modelo multidimensional.
5. Preencher os campos:
    a. Name: quantidade;
    b. Aggregator: sum;
    c. Collum: quantidade; (Nome da métrica que se encontra na tabela fato no banco de dados)
    d. dataType: Numeric
6. Adicionar uma dimensão.
7. Indicar qual a respectiva tabela para esta hierarquia clicando com o botão direito sobre a hierarquia e depois em Add Table. Adicionar as informações da tabela de dimensão correspondente a que foi criada no banco de dados.
8. Adicionar um level para a Hierarquia criada. Esse level é equivalente à um atributo da tabela dimensão que deverá ser hierarquizado na dimensão de modo à serem gerados o *drill down* e *drill up*. Podem existir um ou mais levels em uma tabela dimensão.
9. Preencher os campos
10. Repetir os passo acima para as outras dimensões.

O *schema* gera um arquivo XML que será publicado no Mondrian/Pentaho após comando do usuário. O arquivo XML gerado no trabalho é apresentado a seguir:

```
<Schema name="POC">
  <Cube name="Temperatura" cache="true" enabled="true">
    <Table name="fato_temperatura" />
    <Dimension type="StandardDimension" foreignKey="ID_DIM_COLONIA" name="colonia">
      <Hierarchy hasAll="true" primaryKey="ID_DIM_COLONIA">
        <Table name="dim_colonia" />
        <Level name="numero" column="NUMERO" uniqueMembers="true" hideMemberIf="Never" />
      </Hierarchy>
    </Dimension>
```

```
    <Dimension type="StandardDimension" foreignKey="ID_DIM_ESPECIE" name="especie">
      <Hierarchy hasAll="true" primaryKey="ID_DIM_ESPECIE">
        <Table name="dim_especie" />
        <Level name="Nome cientifico" column="NOME_CIENTIFICO" uniqueMembers="false" />
      </Hierarchy>
    </Dimension>
    <Dimension type="StandardDimension" foreignKey="ID_DIM_TURNO" name="turno">
      <Hierarchy hasAll="true" primaryKey="ID_DIM_TURNO">
        <Table name="dim_turno" />
        <Level name="Turno" column="ID_DIM_TURNO" uniqueMembers="false" />
      </Hierarchy>
    </Dimension>
    <Dimension type="StandardDimension" foreignKey="ID_DIM_LOCALIZACAO" name="localidade">
      <Hierarchy hasAll="true" primaryKey="ID_DIM_LOCALIDADE">
        <Table name="dim_localidade" />
        <Level name="estado" column="ESTADO" uniqueMembers="false" hideMemberIf="Never" />
        <Level name="cidade" column="CIDADE" uniqueMembers="false" hideMemberIf="Never" />
      </Hierarchy>
    </Dimension>
    <Dimension type="StandardDimension" foreignKey="ID_DIM_TEMPO" name="tempo">
      <Hierarchy hasAll="true" primaryKey="ID_DIM_TEMPO">
        <Table name="dim_tempo" />
        <Level name="ano" column="ANO" uniqueMembers="false" hideMemberIf="Never" />
        <Level name="mes" column="MES" uniqueMembers="false" hideMemberIf="Never" />
        <Level name="dia" column="DIA" uniqueMembers="false" hideMemberIf="Never" />
      </Hierarchy>
    </Dimension>
    <Measure  name="TEMP_ENTRADA"  column="TEMP_ENTRADA"  datatype="Numeric"  aggregator="avg"
caption="Temp. entrada" visible="true" />
    <Measure   name="TEMP_CRIA"   column="TEMP_CRIA"   datatype="Numeric"   aggregator="avg"
caption="Temp. cria" visible="true" />
    <Measure  name="TEMP_CONTROLE"  column="TEMP_CONTROLE"  datatype="Numeric"  aggregator="avg"
caption="Temp. controle" visible="true" />
    <Measure  name="TEMP_AMBIENTE"  column="TEMP_AMBIENTE"  datatype="Numeric"  aggregator="avg"
caption="Temp. ambiente" visible="true" />
  </Cube>
</Schema>
```

Esse exemplo representa como funcionará a navegação no cubo OLAP, baseada em dimensões, com hierarquias e níveis, com o objetivo de apresentar as medidas (*Measures*) de acordo com a combinação definida.

Após essa etapa, resta apenas publica o *schema* para visualização no navegador. Para isso deve-se ir em File -> *Publish*.

A visão inicial apresenta os indicadores de temperatura para um critério de filtro: Colônia > Espécie > Turno > Localidade > Mês. A Figura 26 apresenta o

ambiente Mondrian instalado, configurado e acessando dados do modelo dimensional criado neste trabalho.

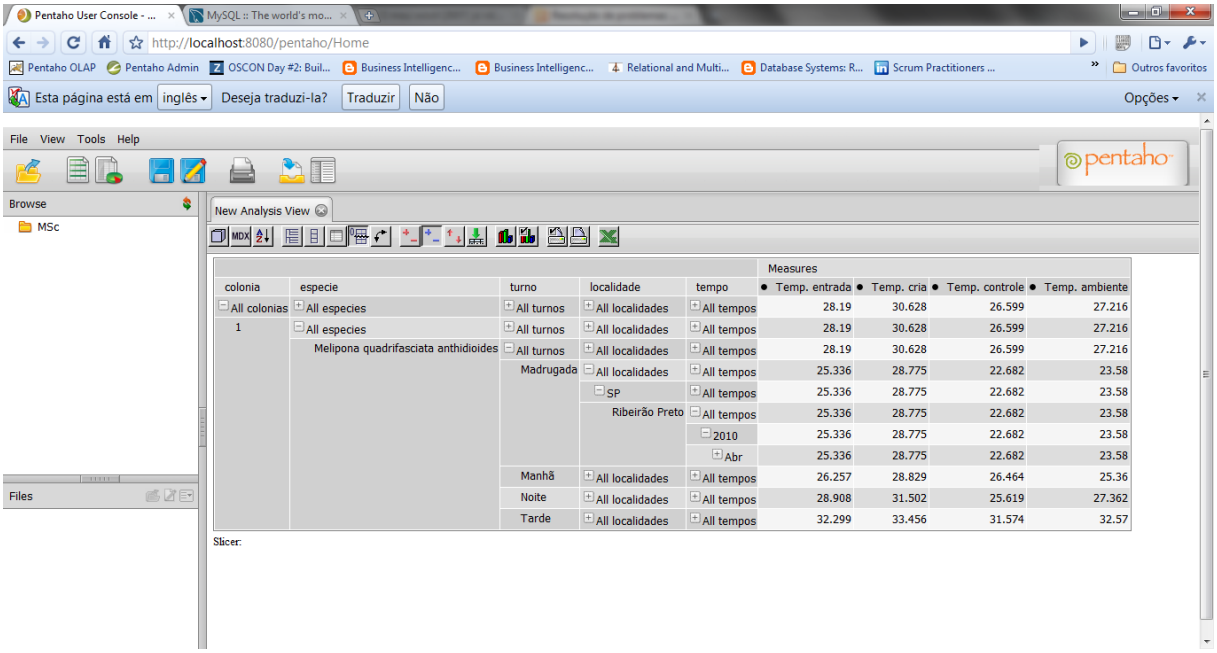


Figura 26. Tela de análise do Mondrian/Pentaho.

Para customizações na análise utiliza-se a barra de ferramentas acima da tabela (Figura 27).

Figura 27. Barra de ferramentas do Mondrian/Pentaho

Dessa forma, o ambiente Mondrian/Pentaho está configurado e funcional, além de preparado para receber novos cubos ou ainda customizar os existentes, e, pronto para auxiliar os tomadores de decisões a extrair dados dos cubos por meio do cruzamento das dimensões existentes.

APÊNDICE B – CODIGOS SQL

Este Apêndice tem como objetivo apresentar os códigos SQL utilizados na criação da prova de conceito.

B.1 INSERT NA TABELA FATO

```
INSERT INTO FATO_TEMPERATURA (ID_DIM_COLONIA, ID_DIM_ESPECIE,
ID_DIM_LOCALIZACAO, TEMP_ENTRADA, TEMP_CRIA, TEMP_CONTROLE,
TEMP_AMBIENTE, ID_DIM_TEMPO, ID_DIM_TURNO)
SELECT
    1 ID_DIM_COLONIA, #ID_COLONIA
    1 ID_DIM_ESPECIE, #ID_DIM_ESPECIE
    2 ID_DIM_LOCALIZACAO, #ID_DIM_LOCALIZACAO
    AVG(TEMP_ENTRADA),AVG(TEMP_CRIA),AVG(TEMP_AMBIENTE),
AVG(TEMP_CONTROLE),
    DT.ID_DIM_TEMPO,
    TU.ID_DIM_TURNO
FROM
    DIM_TEMPO DT, TEMP T, DIM_TURNO TU
WHERE
    DT.ANO = CAST(SUBSTRING(T.data,1,5) AS UNSIGNED)
AND
    DT.MES = 'Abr'
AND
    DT.DIA = CAST(SUBSTRING(data,10,2) as UNSIGNED)
AND
    DT.TURNO = T.TURNO
AND
    T.TURNO = TU.TURNO
GROUP BY
    ANO, MES, DIA, T.TURNO);
```

Figura 28. Trecho de código do ETL

B.2 INSERT NA DIMENSÃO TEMPO

```
INSERT INTO DIM_TEMPO (ANO, MES, DIA, TURNO)
SELECT
    #data,
    SUBSTRING(data,1,5) ANO,
    IF(STRCMP(SUBSTRING(data,7,2),'04'),'Abr',null) MES,
    CAST(SUBSTRING(data,10,2) as UNSIGNED) DIA,
    IF (
        cast(SUBSTRING(data,13,2)    as    UNSIGNED)    >=    0    AND
cast(SUBSTRING(data,13,2) as UNSIGNED) < 6,'Madrugada',
    IF (
        cast(SUBSTRING(data,13,2)    as    UNSIGNED)    >=    6    AND
cast(SUBSTRING(data,13,2) as UNSIGNED) < 12,'Manhã',
    IF (
        cast(SUBSTRING(data,13,2)    as    UNSIGNED)    >=    12    AND
cast(SUBSTRING(data,13,2) as UNSIGNED) < 18,'Tarde',
    IF (
        cast(SUBSTRING(data,13,2)    as    UNSIGNED)    >=    18    AND
cast(SUBSTRING(data,13,2) as UNSIGNED) < 24,'Noite',null)))) TURNO
FROM
    TEMP
GROUP BY ANO, MES, DIA, TURNO;
```

Figura 29. Trecho de código usado para Transformação dos dados.