

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
COORDENAÇÃO DE INFORMÁTICA
CURSO DE TECNOLOGIA EM ANÁLISE E DESENVOLVIMENTO DE SISTEMAS

EMANOÉLY FERNANDA GURA
LARISSA LOURENÇO NUNES BENCK

**CONSTRUÇÃO DE UM *DATA WAREHOUSE*, ALIADO A UMA
FERRAMENTA *OPEN SOURCE IREPORT* NA GERAÇÃO DE
INFORMAÇÕES PARA TOMADA DE DECISÃO**

TRABALHO DE CONCLUSÃO DE CURSO

PONTA GROSSA

2011

EMANOÉLY FERNANDA GURA
LARISSA LOURENÇO NUNES BENCK

**CONSTRUÇÃO DE UM *DATA WAREHOUSE*, ALIADO A UMA
FERRAMENTA *OPEN SOURCE IREPORT* NA GERAÇÃO DE
INFORMAÇÕES PARA TOMADA DE DECISÃO**

Trabalho de Conclusão de Curso de graduação, apresentado à disciplina de Trabalho de Diplomação, do Curso Superior de Tecnologia em Análise e Desenvolvimento de Sistemas – COADS – da Universidade Tecnológica Federal do Paraná – UTFPR, como requisito parcial para obtenção do título de Tecnólogo.

Orientadora: Prof. Msc. Simone de Almeida

PONTA GROSSA
2011



TERMO DE APROVAÇÃO

CONSTRUÇÃO DE UM *DATA WAREHOUSE*, ALIADO A UMA FERRAMENTA *OPEN SOURCE IREPORT* NA GERAÇÃO DE INFORMAÇÕES PARA TOMADA DE DECISÃO

por

EMANOÉLY FERNANDA GURA
LARISSA LOURENÇO NUNES BENCK

Este Trabalho de Conclusão de Curso (TCC) ou esta Monografia ou esta Dissertação foi apresentado(a) em 16 de novembro de 2011 como requisito parcial para a obtenção do título de Tecnólogo em Análise e Desenvolvimento de Sistemas. As candidatas foram argüidas pela Banca Examinadora composta pelos professores abaixo assinados. Após deliberação, a Banca Examinadora considerou o trabalho aprovado.

Prof. Msc. Simone de Almeida
Prof. Orientador

Prof. Marcos Vinicius Fidelis
Membro titular

Prof. Dr. Saulo Jorge Beltrão de Queiroz
Membro titular

Prof. Msc. Helyane Bronoski Borges
Responsável pelos Trabalhos
de Conclusão de Curso

Prof. Dr. André Koscianski
Coordenador do Curso
UTFPR - Campus Ponta Grossa

AGRADECIMENTOS

Este Trabalho de Conclusão de Curso não seria possível sem a oportunidade nos dada por Deus de ingressar na UTFPR, uma universidade de tamanho prestígio e conceito. Agradecemos acima de tudo Àquele que tanto nos proporcionou durante esses três anos de curso.

Um agradecimento especial à professora e amiga Simone de Almeida que tanto nos auxiliou como alunas, quanto como orientandas, que, confiando em nós, instigou o tema do trabalho, dando-nos força para que ele pudesse ser concluído. Também agradecemos sua disponibilidade, compreensão e paciência durante o desenvolvimento do trabalho.

Agradecemos também, a todos os professores que nos incentivaram, apoiaram e inspiraram no decorrer do curso.

Aos nossos amigos e colegas de universidade, em especial à Jessyca, que nos incentivou a acordar cedo todos os dias para não perder sua carona.

Os agradecimentos específicos de Emanoély

Sinto-me grata a todos que me apoiaram e incentivaram, seja através de palavras motivadoras ou com auxílios e companheirismo para a realização desse projeto. Agradeço, de forma singular, aos meus pais Eutêmia e Estanislau, que não só me consolaram em momentos de dificuldade, como foram meus maiores pilares de sustentação durante todo o curso e em especial, durante a realização do Trabalho de Diplomação. Ao meu irmão Evandro, que devo agradecimentos não só pelos conselhos e apoio a mim dispensado, mas por toda a ajuda dedicada nos meus projetos do curso.

Agradeço a minha companheira de trabalho e de curso, Larissa, que se mostrou mais que uma colega, se tornando uma amiga e irmã, que me aturou durante esses três anos, mas principalmente nos últimos três meses. Agradeço pela sua dedicação e esforço e por me dar a oportunidade de realizar o trabalho em conjunto.

Em especial, agradeço a todos os professores da coordenação de informática, que além de aulas, lecionaram lições e conselhos. Em particular, ao meu chefe e professor, Marcos Vinícius Fidelis, que me permitiu algumas fugidas do trabalho para poder finalizar o trabalho de diplomação.

Também, de nada valeria tanta dedicação, se não pudesse comemorar todo o esforço com os meus amigos que tanto me apoiaram e entenderam minha falta de tempo e paciência. Agradeço especialmente a Ariane, que apesar das broncas, se tornou minha irmã para todo o

sempre; àquelas que não perdiam a “janta das meninas” lá em casa; àqueles que me puxavam pro boliche pra descontar a raiva; àquele que aturou minha falta de paciência nos últimos dias; aos meus amigos de Irati, que talvez ainda não entendam o motivo do meu sumiço e àqueles que eu levarei pra sempre em minha história.

Os agradecimentos específicos de Larissa

Agradeço em especial todas as pessoas envolvidas com minha formação, desde apoio motivacional, incentivador, esperançoso e financeiro dos meus pais e avós para realizar um curso fora de minha cidade; principalmente nas minhas primeiras semanas do curso, na fase de adaptação eles sempre estavam lá para me ajudar, não se esquecendo da minha grande amiga Amanda que desde o terceiro ano consolidou-se nossa amizade e veio lapidando-se nesses 3 anos de convivência; morando junto, estudando na mesma faculdade, respeitando os limites de cada uma e por fim nos ajudando a morar longe dos pais. Não se esquecendo das minhas irmãs Letícia e Luiza que sempre estavam contando-me piadinhas e dando risadas, nos momentos mais ansiosos e difíceis do trabalho, fazendo-me relaxar, restaurar forças para continuar.

Dentro da faculdade, não posso esquecer-me de dois grandes amigos que sempre me apoiaram e estavam comigo para vários momentos, o Leonardo e a Emanoély, que acabou se tornando uma nova grande amiga para conversas, risadas, boliches e estudos principalmente; madrugadas estudando para provas e desenvolvendo projetos, e o mais engraçado era quando estávamos com sono e confundíamos as palavras como escalabilidade com “escanolabilidade”, agradeço muito por ter concedido essa oportunidade de ser sua amiga e desenvolvermos nosso trabalho de diplomação juntas.

Agradeço todos os professores da Coordenação de Informática, em especial a professora Simone Nasser que me incentivou no curso, e concedeu uma vaga ao grupo de iniciação científica em que obtive muitas experiências.

"Nunca deixe que alguém te diga que não pode fazer algo. Nem mesmo eu. Se você tem um sonho, tem que protegê-lo. As pessoas que não podem fazer por si mesmas, dirão que você não consegue. Se quer alguma coisa, vá e lute por ela. Ponto final."

(Chris Gardner)

RESUMO

GURA, Emanóely F; BENCK, Larissa L N. **Construção de um *Data Warehouse*, aliado a uma ferramenta *open source iReport* na geração de informações para tomada de decisão.** 2011. 89f. Trabalho de Conclusão de Curso em Tecnologia em Análise e Desenvolvimento de Sistemas – Universidade Tecnológica Federal do Paraná. Ponta Grossa, 2011.

Na maioria das organizações, é fato a problemática de falta da correta interpretação das informações e a dificuldade de agrupá-las, para obter um novo conhecimento. Para tal solução a literatura apresenta a tecnologia de *Data Warehouse*. Esta tecnologia é uma solução que procura de maneira flexível e eficiente tratar um grande volume de dados e obter informações que auxiliem no processo de tomada de decisão. Desta forma, este trabalho descreve o processo de construção de uma aplicação que explora a tecnologia de *Data Warehouse*. Os dados foram coletados em uma farmácia magistral localizada na região de Ponta Grossa, e padronizados para facilitar a consulta e extração de conhecimento. Para implementação do estudo de caso foi utilizada a ferramenta *Pentaho*, que oferece suporte para todo o processo do *Data Warehouse* e que mostrou-se adequada no estudo de caso. Apresenta-se também a ferramenta *iReport*, responsável pela geração dos relatórios e que ao final do estudo apresentou integração total com a *Pentaho*, gerando os relatórios, conforme solicitados. Além disso, também são elencadas as dificuldades e benefícios encontrados com o desenvolvimento do trabalho.

Palavras-chave: *Data Warehouse*. *iReport*. *Pentaho*. Gestão da Informação. *Business Intelligence*.

ABSTRACT

GURA, Emanóely F. BENCK; Larissa L N. **Construction of a Data Warehouse, coupled with open source tool iReport in generating information for decision making.** 2011. 89f. Completion of Course Work in Technology Analysis and Systems Development – Federal Technological University of Parana. Ponta Grossa, 2011.

In most business organizations, it is a fact the problem of lack of the correct interpretation of information and difficulties of grouping them, for getting a new knowledge. To this solution presents the literature data warehouse technology. This technology is a solution that looks for a flexible and efficient options to treat a huge volume of data and obtain information that help in the decision-making process. Thus, this work presents process of constructing an application that discovery the Data Warehouse technology. The data has been collected in a the pharmacy located in the Ponta Grossa region, and standardized to facilitate the consult and extraction of knowledge. For implementing the case study, it was used the *Pentaho* tool, that supports the entire process of *Data Warehouse* and proved adequate in the case study. It was showed the iReport tool, responsible for generating reports and the end of the study had full integration with Pentaho, generating reports, according to requested. In addition, also the difficulty and benefic discovered are listed with the development of the work.

Keywords: Data Warehouse. iReport. Pentaho. Information management. Business Intelligence.

LISTA DE QUADROS

Quadro 1 - Diferença entre Banco de Dados Operacionais e Analíticos	22
Quadro 2 - Diferenças OLTP e OLAP.....	23
Quadro 3 - Doze Regras do Ambiente OLAP	34
Quadro 4 - Diferenças entre <i>Pentaho</i> e <i>SQL Server 2008</i>	39
Quadro 5 – Explicação do comando SQL utilizado na transformação medico.....	51

LISTA DE FIGURAS

Figura 1 - Integração de dados	24
Figura 2 - Diagrama de Entidade e Relacionamento	26
Figura 3 - Herança	26
Figura 4 - Agregação.....	27
Figura 5 - Modelo Estrela.....	29
Figura 6 - Ligação dos Usuários aos metadados	30
Figura 7 - Processo de DataWarehousing	31
Figura 8 - <i>Pentaho Schema Workbench</i>	38
Figura 9 - Ciclo de vida do <i>Data Warehouse</i>	42
Figura 10 - Modelagem Multidimensional.....	43
Figura 11 - Diagrama do processo de DW.....	45
Figura 12 - Configuração do BD	46
Figura 13 - Conexão com o banco de dados	46
Figura 14 - Conexão com o BD.....	47
Figura 15 - Modelo <i>Schema</i>	47
Figura 16 - XML do Cubo.....	48
Figura 17 - Publicação do XML.....	48
Figura 18 - Conexão do banco na <i>Data Integration</i>	49
Figura 19 - Transformação cliente.....	50
Figura 20 – Mensagem de sucesso da transformação cliente.....	50
Figura 21 - Transformação médico e especialidade juntamente com a mensagem de sucesso	52
Figura 22 - Transformação Venda e mensagem de sucesso.....	53
Figura 23 - Transformação Componente e mensagem de sucesso	54
Figura 24 - Transformação Produto.....	54
Figura 25 - Transformação Tempo	55
Figura 26 - Transformação Fato venda	56
Figure 27 - Ambiente de desenvolvimento de relatórios da ferramenta <i>Pentaho</i>	59
Figura 28 - Menu Arquivo.....	61
Figura 29 - <i>New File</i>	62
Figura 30 - Local para salvar o relatório.....	62

Figura 31 - <i>Template</i> Principal	63
Figura 32 - Exclusão de título	64
Figura 33 - <i>Template</i> Inicial mostrando linhas excluídas	64
Figura 34 - Palheta	65
Figura 35 - Adicionar componente	65
Figura 36 - Adicionar parâmetros.....	67
Figura 37 - Editar <i>Query</i>	68
Figura 38 - Editar <i>Query</i>	68
Figura 39 - Submenu <i>Fields</i>	69
Figura 40 - <i>Header</i>	70
Figura 41 - <i>Preview</i>	70
Figura 42 - Relatório	70
Figura 43 - Relatório Médico Faturamento	71
Figura 44- Relatório Faturamento Cliente	72
Figura 45 - Relatório Faturamento Componente	73
Figura 46 - Relatório médico por especialidade	74
Figura 47 - Relatório maior e menor faturamento do ano 2011	75
Figura 48- Relatório total de faturamento para os anos 2007 a 2011	75
Figura 49- Instalação do <i>iReport</i>	82
Figura 50 – Instalação do <i>iReport</i>	82
Figura 51- Instalação do <i>iReport</i>	83
Figura 52 - Instalação do <i>iReport</i>	83
Figura 53 - Instalando o <i>iReport</i>	84
Figura 54 - Processo de Instalação Completo	84
Figura 55 - Tela Inicial do <i>iReport</i>	85
Figura 56 - Report Datasources	85
Figura 57 - Criação da nova conexão	86
Figura 58 - Escolha do tipo de <i>Datasource</i>	86
Figura 59 - Conexão com o <i>database connection</i>	87
Figura 60 - Conexão com Sucesso.....	87
Figura 61 – Modelo de entidade e relacionamento.....	88

LISTA DE SIGLAS

BD	Banco de Dados
BI	<i>Business Intelligence</i> (Inteligência Empresarial)
DM	<i>Data Mart</i> (Dados Departamentais)
DW	<i>Data Warehouse</i> (Armazém de Dados)
ER	Entidade e Relacionamento
ETL	<i>Extract, transformation e load</i> (Extração, transformação e carga)
OLTP	Processamento de Transações <i>Online</i>
OLAP	Processamento Analítico <i>Online</i>
SQL	<i>Structure Query Language</i> (Linguagem de Estruturação de Dados)

SUMÁRIO

1 INTRODUÇÃO	15
1.1 OBJETIVOS	15
1.1.1 Objetivo Geral	16
1.1.2 Objetivos Específicos.....	16
1.2 PROBLEMA.....	16
1.3 ORGANIZAÇÃO DO TRABALHO	17
2 DATA WAREHOUSE	18
2.1 DADOS, INFORMAÇÃO E CONHECIMENTO	18
2.2 DIFERENÇAS ENTRE GESTÃO DO CONHECIMENTO E da INFORMAÇÃO	19
2.3 TOMADA DE DECISÃO NO SISTEMA DE INFORMAÇÃO	20
2.4 DEFINIÇÃO DE <i>DATA WAREHOUSE</i>	21
2.5 CARACTERÍSTICAS DO <i>DATA WAREHOUSE</i>	23
2.5.1 Orientado a Assunto.....	23
2.5.2 Variação de Tempo	23
2.5.3 Não Volátil	24
2.5.4 Integração	24
2.6 Modelagem dos dados.....	25
2.6.1 Modelagem Entidade-Relacionamento (ER).....	25
2.6.2 Modelagem Multidimensional.....	27
2.7 METADADOS	29
2.8 <i>DATA MART</i>	30
2.9 GRANULARIDADE	30
2.10PROCESSO DE <i>DATA WAREHOUSING</i>	31
2.11TIPOS DE IMPLEMENTAÇÕES.....	32
3 FERRAMENTAS DE <i>DATA WAREHOUSE</i>	33
3.1 APRESENTAÇÃO	33
3.2 <i>MICROSOFT® SQL SERVER 2008 EXPRESS</i>	34
3.2.1 Componentes	35
3.2.2 Caso de Sucesso.....	36
3.3 <i>PENTAHO®</i>	37
3.3.1 Componentes	37
3.3.2 Caso de Sucesso.....	37
3.4 COMPARAÇÃO	38
4 ESTUDO DE CASO	40
4.1 DESCRIÇÃO.....	40
4.2 PRÉ-PROCESSO.....	41
4.3 PROJETO LÓGICO.....	42
4.4 FERRAMENTAS E SUAS CONFIGURAÇÕES	43
4.5 DESENVOLVIMENTO DO ESTUDO DE CASO	44
4.5.1 Atividade 1 - Criação da base de dados	44
4.5.2 Atividade 2 - Gerar XML do modelo multidimensional.....	44
4.5.3 Atividade 3 - Extrair, transformar e carregar a base de dados	49
4.6 POTENCIAL DE EXPANSÃO.....	56
5 IREPORT	58
5.1 IREPORT E PENTAHO	58
5.2 DEFINIÇÕES	59

5.2.1 <i>JasperReports</i>	59
5.2.2 JDK	60
5.2.3 <i>iReport</i>	60
5.2.4 <i>Jar</i>	60
5.3 REQUISITOS PARA DESENVOLVIMENTO	61
5.4 O AMBIENTE <i>IREPORT</i>	61
5.4.1 Criando o <i>Template</i>	61
5.4.2 Conhecendo os componentes	64
5.4.3 A <i>Aba Tools</i>	66
5.4.4 Palheta de Componentes	66
5.4.5 Modelando o Relatório.....	67
5.5 IMPLEMENTAÇÃO DO ESTUDO DE CASO	71
6 CONCLUSÃO	76
7 TRABALHOS FUTUROS	78
REFERÊNCIAS.....	79
APÊNDICE A - INSTALAÇÃO E CONFIGURAÇÃO DO <i>IREPORT</i>.....	82
ANEXO B – DECLARAÇÃO DE CIENCIA.....	89

1 INTRODUÇÃO

Polêmicas sempre foram originadas sobre quais pessoas seriam capazes de interpretar e compreender as palavras; interpretar não somente em um contexto qualquer e sim em vários outros contextos. Quem desenvolve essa capacidade percebe-se que obtém uma grande vantagem sobre suas atitudes e comportamento. Isso não é diferente em uma organização, que nada mais é do que uma junção de pessoas e funções que gira totalmente em torno de informações. É indiscutível que a informação é considerada a melhor forma de negociação atualmente, pois é um requisito relevante para manter a vantagem competitiva, auxiliando no processo de tomada de decisão.

Hoje as empresas não trabalham mais para gerar e implantar informações nos sistemas, mas sim, para obter informações gerenciais executivas para melhor agir e mais rápido decidir. Faz-se necessário organizar, controlar e filtrar os dados para melhor apoiar os executivos, gerentes e analistas, surgindo, então, a área de *Business Intelligence*.

Como a maioria dos sistemas não são modelados e preparados para armazenar informações decisórias ou estratégicas nas empresas, surge à necessidade de uma base histórica para as tomadas de decisões, denominado *Data Warehouse* (DW) ou Armazém de Dados. O *Data Warehouse* integra os dados internos e externos de uma organização, proporcionando assim, a conectividade do mundo virtual com o dos negócios.

Com esse agrupamento de dados é possível gerar relatórios com novas visões, ou seja, a abertura de um leque de informações obscuras até então, transformando o conhecimento antes tácito em tangível.

Este trabalho apresenta uma revisão bibliográfica sobre *Data Warehouse*, mostrando diferenças e conceitos sobre os dados, conhecimento e informação. Foca no que é o *Data Warehouse*, suas características e como implementá-lo. Também traz uma análise de duas ferramentas disponíveis no mercado, para auxiliar na decisão de qual se adéqua ao estudo de caso que é desenvolvido. Por fim, traz o conhecimento de uma ferramenta *open source* – *iReport* – para análise de dados, que pode ser considerada a última etapa do processo de *Data Warehouse*.

1.1 OBJETIVOS

Os objetivos deste trabalho estão descritos a seguir.

1.1.1 Objetivo Geral

Conhecer o processo de *Data Warehouse* e aplicá-lo em um estudo de caso, passando por todas as suas fases e utilizando todos os conceitos e princípios estudados, para geração de conhecimento. Pessoas menos experientes com o estudo podem conhecer e ter um melhor entendimento sobre seu uso e aplicabilidade.

1.1.2 Objetivos Específicos

- Descrever sobre gestão da informação, estudar e compreender *Data Warehouse*.
- Pesquisar ferramentas e formas de desenvolvimento de uma solução utilizando um *Data Warehouse*.
- Aplicar o *iReport* em um estudo de caso, apresentando as vantagens e dificuldades encontradas.
- Discorrer sobre a ferramenta *open source* utilizada para a extração da informação.
- Descrever resultados e dificuldades encontrados com a utilização da ferramenta *open source*.
- Apresentar a transformação do conhecimento, bem como as dificuldades encontradas.
- Utilizar um caso prático para experimentar a utilização da tecnologia de *Data Warehouse* e *iReport*.

1.2 PROBLEMA

Percebe-se a grande dificuldade encontrada pelas organizações na hora de tomar decisões rápidas. São poucas empresas que conseguem manter informações atualizadas e ao alcance dos administradores e gerentes em tempo real. Muitos utilizam-se apenas de sua experiência (ou a falta dela) na hora de decidir e, muitas vezes, são escolhas vitais para a organização, que deveriam ter sido baseadas em um estudo histórico. Além disso, muitas organizações não sabem como trabalhar com os dados coletados ao longo dos anos, perdendo

toda a sua essencialidade e diferenciação no mundo dos negócios. Dessa forma, este trabalho pretende utilizar a tecnologia de *Data Warehouse* associada à ferramenta *iReport* para apoiar a tomada de decisão.

1.3 ORGANIZAÇÃO DO TRABALHO

O estudo sobre *Data Warehouse* foi organizado em seis capítulos. O Capítulo 2 aborda e define, por meio da literatura especializada, o que é um *Data Warehouse*, apresentando suas principais características, trazendo definições dos principais autores da área e definindo alguns conceitos de utilização e desenvolvimento.

O Capítulo 3 discorre sobre a escolha da ferramenta para a realização do *Data Warehouse*. São analisadas duas ferramentas para a implantação de um *Data Warehouse*.

No Capítulo 4 é apresentado o estudo de caso, sua análise e desenvolvimento, sobre o qual foi aplicada a tecnologia de *Data Warehouse*.

O Capítulo 5 apresenta a ferramenta *iReport*, utilizada na geração dos relatórios, como parte do processo de *Data Warehouse*.

O Capítulo 6 finaliza este trabalho apresentando os resultados obtidos na aprendizagem e no desenvolvimento de *Data Warehouse*.

2 DATA WAREHOUSE

Este capítulo descreve conceitos sobre dados, informação e conhecimento, gestão da informação e considerações relacionadas à *Data Warehouse* encontradas na literatura. A Seção 2.1 apresenta as diferenças e conceitos sobre dados, informação e conhecimento. A Seção 2.2 traz algumas diferenças entre gestão do conhecimento e da informação. Na Seção 2.3 é exposto a área da informática que engloba a tomada de decisão. A Seção 2.4 descreve algumas definições propostas por pesquisadores da área sobre *Data Warehouse*. A Seção 2.5 relata as principais características de um *Data Warehouse*. Na seção 2.6 encontram-se definições sobre modelagem de dados e na seção 2.7 definições sobre metadados. A seção 2.8 define *Data Mart* e na seção 2.9, granularidade de dados. O processo de *Data Warehousing* é descrito na seção 2.10 e a seção 2.11 relata sobre os tipos de implementações do *Data Warehouse*.

2.1 DADOS, INFORMAÇÃO E CONHECIMENTO

Muitos se perguntam por que está havendo esse súbito interesse em conhecimento. E duas afirmações colocam discussão ao tema. Primeira: não é um súbito interesse, uma vez que a busca pelo conhecimento vem crescendo consideravelmente, mas hoje o apelo é digital e aguçado em tecnologia; e a segunda é sustentada por Davenport & Prusak (1998, p.15): “Numa economia global, o conhecimento pode ser a maior vantagem competitiva da empresa”.

Pode-se, logicamente, atribuir essa busca acelerada por conhecimento às tecnologias, que dão uma falsa visão de que aqueles que a dispõem sofisticadamente obtêm a melhor informação. A tecnologia ajuda a juntar valores e transformar dados em informação, o que não significa que um montante de dados armazenado será útil. É necessário que eles sejam transformados em informação e esta em conhecimento. Percebe-se então que dados diferem de informação, assim como informação difere-se de conhecimento.

Na definição de Davenport & Prusak (1998, p. 02), “Dados são um conjunto de fatos distintos e objetivos, relativos a eventos. Num contexto organizacional, dados são utilitariamente descritos como registros estruturados de transações.”

Analisando assim, percebe-se que apesar do gigantesco número de dados armazenados nas organizações, eles não passam de espaço ocupado se não forem integrados em um determinado conceito. Davenport & Prusak (1998, p. 03 *apud* Drucker) caracterizam

os dados como sendo dotados de relevância e propósito, o que dá uma dimensão de sua importância se trabalhados da maneira correta. Como citam Davenport & Prusak (1998, p.03), somente dados não fornecem sustentações para julgamentos, interpretações ou bases para a tomada de decisão e ainda complementam afirmando que “Embora a matéria-prima do processo decisório possa incluir dados, eles não podem dizer o que fazer”. Ou seja, os dados são a base, a matéria-prima para a criação de informação, influenciando no julgamento e no comportamento do usuário.

Trabalha-se com os conceitos de dados, informação e conhecimento em escala de grau, isto é, dados são transformados em informação e a informação gera o conhecimento. O conhecimento não é puro nem simples; é uma mistura de vários elementos. Sendo assim ele é mais próximo da ação do que os dados ou a informação, considerando que deve ser avaliado no momento da tomada de decisão.

2.2 DIFERENÇAS ENTRE GESTÃO DO CONHECIMENTO E DA INFORMAÇÃO

Com a grande quantidade de informação dentro de uma organização ultimamente, a melhor maneira de utilizá-la adequadamente é saber interpretá-la, utilizando técnicas adquiridas através da Gestão de Conhecimento e da Gestão da Informação. Mas, afinal qual a diferença entre as duas gestões?

De acordo com Valentim (2004), Gestão de Informação é um conjunto de estratégias que visam identificar as necessidades informacionais, mapear os fluxos formais de informação nos diferentes ambientes da organização, assim como sua coleta, filtragem, análise, organização, objetivando apoiar o desenvolvimento das atividades cotidianas e a tomada de decisão.

Já a Gestão do Conhecimento é o conjunto de estratégias para criar, adquirir, compartilhar e utiliza alvos de conhecimento, bem como estabelece fluxos que garantam a informação necessária no tempo e formato adequados, a fim de auxiliar na geração de ideia, solução de problemas e tomada de decisão.

Na literatura observa-se que algumas correntes fundem os dois modelos de gestão. As duas focam para o fato de que pretendem apoiar as atividades desenvolvidas no dia a dia e na tomada de decisão, mas focam em fluxos informacionais diferenciados, sendo a Gestão de

Informação nos fluxos formais (conhecimento explícito)¹ e a Gestão de Conhecimento nos fluxos informais (conhecimento tácito)².

2.3 TOMADA DE DECISÃO NO SISTEMA DE INFORMAÇÃO

Sabe-se que Sistema de Informação Gerencial (SIG) é um instrumento administrativo de significativo auxílio para os executivos das empresas unindo informação e conhecimento. É ele que auxilia os gerentes e administradores na consolidação da qualidade, produtividade e participação da empresa. Porém, para Oliveira (2005) tem-se dificuldade de avaliar quais seriam os efetivos benefícios de um SIG, isto é, a real melhoria no processo decisório. Eles atuam como elementos polarizadores de eventos empresariais provenientes dos períodos de atividades, sejam elas internas (como controle de estoques, alocação de recursos) ou externas (como interações com fornecedores ou clientes).

O SIG é projetado para oferecer aos administradores informações seguras para a tomada de decisão, sendo esse o ponto de interesse do estudo da informática nas organizações. O conhecimento adquirido através dos Sistemas de Informação é o processo de comunicação mediante as quais são fornecidos os elementos básicos para essas decisões. As tomadas de decisões envolvem um ciclo e é fundamental a existência de informações apropriadas a cada uma das fases do ciclo. Esse ciclo pode ser definido pelas fases de implantação, avaliação da decisão, recomendações de mudanças e tomada de decisão. Cassaro (1999) afirma que o fato de poder contar com informações adequadas e oportunas em cada fase, é de vital importância para o sucesso da empresa e consecutivamente dos administradores.

Pode-se resumir, portanto, que os Sistemas de Informações, na modalidade Gerencial, englobam processos que auxiliam executivos na tomada de decisão baseada na informação. Cabe aos analistas e desenvolvedores de tais sistemas adaptá-los e implementá-los quanto aos pontos inerentes ao cenário analisado. Essa área de estratégia empresarial é conhecida como *Business Intelligence*, ou Negócio Inteligente, que tem por finalidade auxiliar as organizações através da automatização dos processos empresariais. Faz parte dessa área os processos, entre outros, de *Data Warehouse* ou Armazém de dados, e *Data Mining* ou Mineração de Dados.

¹ Explícito, do latim *explicitus*, quando o conhecimento está declarado, mostrado, explicado, através de papéis, disquetes, cd-rom.

² Tácito, do latim *tacitus*, quando o conhecimento não pode ser exteriorizado por palavras; como: reuniões, eventos, construção individual de conhecimento, valores, crenças e comportamento organizacional.

2.4 DEFINIÇÃO DE *DATA WAREHOUSE*

Pode-se encontrar na literatura algumas definições sobre *Data Warehouse* (DW). Para Oliveira (2002, p.01) “O *Data Warehouse* consiste em organizar os dados corporativos de maneira integrada, com uma única versão da verdade, histórico, variável com o tempo e gerando uma única fonte de dados”.

Machado (2004, p. 22) afirma que:

Data Warehouse é uma arquitetura, não uma tecnologia. Representa uma grande base de dados capaz de integrar, de forma concisa e confiável, as informações de interesse para a empresa, que se encontram espalhadas pelos sistemas operacionais e em fontes externas, para posterior utilização nos sistemas de apoio à decisão.

Tomando por base os conceitos descritos, pode-se perceber que a preocupação com a transformação de dados em informação é acentuada e crescente. Com o passar do tempo houve um aumento considerável nos sistemas de gestão das mais variadas organizações e como consequência o armazenamento e tratamento de dados também cresceu. Para analisar todo esse volume de dados surge o *Data Warehouse*, ou Armazém de Dados, um conceito que vai muito além de um simples repositório. Por mais que pareça simples, a tarefa de organizar esses dados deve ser muito cuidadosa, a seleção e otimização deve ser apurada já que a prioridade no ambiente de DW é o processamento de consultas e não o de transações.

Segundo Oliveira (2002) “o valor do *Data Warehouse* não está em colecionar dados e sim em saber gerenciar aqueles dados sendo transformados em informações úteis”. Machado (2004) complementa, afirmando que algumas das justificativas para a aplicação de tecnologia de *Data Warehouse* em uma organização é a existência na empresa de:

- Várias plataformas de *hardware* e de *software*;
- Constantes alterações nos sistemas transacionais corporativos;
- Dificuldade acentuada na recuperação de dados históricos em períodos superiores ao ano atual de operações;
- Existência de sistemas “pacotes” de fornecedores diferentes;
- Falta de padronização e integração dos dados existentes nos diversos sistemas;
- Carência de documentação e segurança no armazenamento dos dados.

Nesse momento, cabe ressaltar as diferenças encontradas entre os sistemas de banco de dados. Existem os bancos de dados operacionais e os bancos de dados analíticos. Bancos de dados operacionais auxiliam em todas as operações de uma organização, sendo utilizado com frequência para registrar a situação da organização. Quem geralmente manipula e

trabalha com esse tipo de base de dados, são os desenvolvedores do projeto, sejam eles analistas ou programadores. Os bancos de dados analíticos registram a história da organização, dando suporte aos administradores e gerentes no momento de tomada de decisão e sendo pouco utilizados. Pode-se observar no Quadro 1, uma comparação entre esses bancos de dados.

Banco de Dados Operacionais	Banco de Dados Analítico
Segurança	Velocidade no acesso
Integridade	Sumarizado
Normalização	Desnormalizado
Atualização constante	Não atualizados constantemente

Quadro 1 - Diferença entre Banco de Dados Operacionais e Analíticos
Fonte: Autoria Própria.

O DW visa sumarizar os dados, ou seja, trabalha com banco de dados analítico, em que o processamento analítico examina os aspectos dos dados buscando encontrar tendências ou relações entre os dados. Para descrever o modo de processamento nos sistemas de bancos de dados, usam-se os termos OLTP (*Online Transaction Processing* ou Processamento de Transações *Online*) e OLAP (*Online Analytical Processing* ou Processamento Analítico *Online*).

Sistemas baseados em OLTP necessitam processar milhões de transações diariamente, com consistência e capacidade de recuperação de dados, com necessidade de minimizar a concorrência a esses dados e atualização constante.

Já os sistemas de suporte a decisão, baseados em OLAP requerem um histórico de transações em períodos de tempo, com dados estáticos e completos. Esses sistemas processam uma transação por ciclo de transferência.

O Quadro 2 apoia as definições citadas e demonstra mais claramente as diferenças observadas entre os dois tipos de sistemas.

Dados Primitivos/ Dados Operacionais –OLTP	Dados Derivados/Dados Sistemas de Apoio a Decisão - OLAP
Baseados em aplicações	Baseados em assuntos ou negócios
Detalhados	Resumidos ou refinados
Exatos em relação ao momento de acesso	Representam valores de momentos já decorridos ou instantâneos (histórico)
Atendem a comunidade funcional	Atende a comunidade gerencial
Podem ser atualizados	Não são atualizados
São processados repetitivamente	Processados de forma heurística
Requisitos de processamento conhecidos com	Requisitos de processamento não são conhecidos

antecedência	com antecedência
A performance é fundamental	Performance para análise
Acessados uma unidade por vez	Acessados um conjunto por vez
Voltados para transações	Voltados para análise
O controle de atualizações é atribuído de quem tem a posse	O controle de atualização não é problema
Alta disponibilidade	Disponibilidade atenuada
Gerenciados em sua totalidade	Gerenciados por subconjuntos
Não contemplam a redundância	A redundância é permitida
Estrutura fixa	Estrutura flexível
Pequena quantidade de dados usados em um processo	Grande quantidade de dados usados em um processo
Atende às necessidades cotidianas	Atende às necessidades gerenciais
Alta probabilidade de acesso	Baixa ou modesta probabilidade de acesso

Quadro 2 - Diferenças OLTP e OLAP

Fonte: Machado (2004, p. 16).

2.5 CARACTERÍSTICAS DO DATA WAREHOUSE

Além de compreender o que é um *Data Warehouse*, para a sua construção deve-se entender algumas características relevantes. Segundo Inmon (1999, p.14), “*Data Warehouse* é uma coleção de dados orientados por assuntos, integrados, variáveis com o tempo e não voláteis, para dar suporte ao processo de tomada de decisão”. Explica-se em mais detalhes a seguir.

2.5.1 Orientado a Assunto

O DW é orientado ao redor do principal assunto da organização ou de um determinado processo relevante para essa organização e só utilizará os dados que serão usados no processamento do sistema de apoio a decisão. Seja um processo de venda ou de locação, por exemplo, é o processo que movimenta a empresa e esse será o assunto escolhido. É a escolha correta e precisa da necessidade essencial do negócio que garantirá ou não o sucesso do projeto.

2.5.2 Variação de Tempo

Necessita de dados precisos em relação ao tempo. Esses dados são denominados *snapshots*, ou foto instantânea. Caso o *snapshot* não esteja correto, outra “foto” poderá ser tirada, mas assim que se assume que os dados estão corretos, eles não sofrem mais alterações. São registros estáticos capturados num momento de tempo pré-determinado, em que a variável tempo nunca muda, ou seja, é um dado que não pode ser atualizado.

2.5.3 Não Volátil

Os dados carregados no DW não podem mais ser alterados. Isso significa que o DW tem duas operações: a carga dos dados e o acesso a esses dados (consultas), portanto nada é alterado como em um Banco de Dados Transacional (onde são realizadas inclusões, exclusões, alterações e consultas diárias).

2.5.4 Integração

Inmon (1999) classifica a integração como a mais importante característica dos sistemas de DW. Os dados necessitam ser armazenados de forma única, quando são movidos do ambiente operacional para o ambiente analítico. A migração de dados de diferentes bases remotas para a base principal deverá depender de uma integração dos dados, convencionando nomes, variáveis de medidas e assim sucessivamente, como pode ser exemplificado na Figura 1. Neste exemplo, escolheu-se aleatoriamente a padronização da aplicação A, adotando os valores M, F para a variável sexo e DDMMAA para a variável data. Sendo assim, os valores das outras aplicações passaram pelo processo de transformação para que os dados do DW estejam integrados.

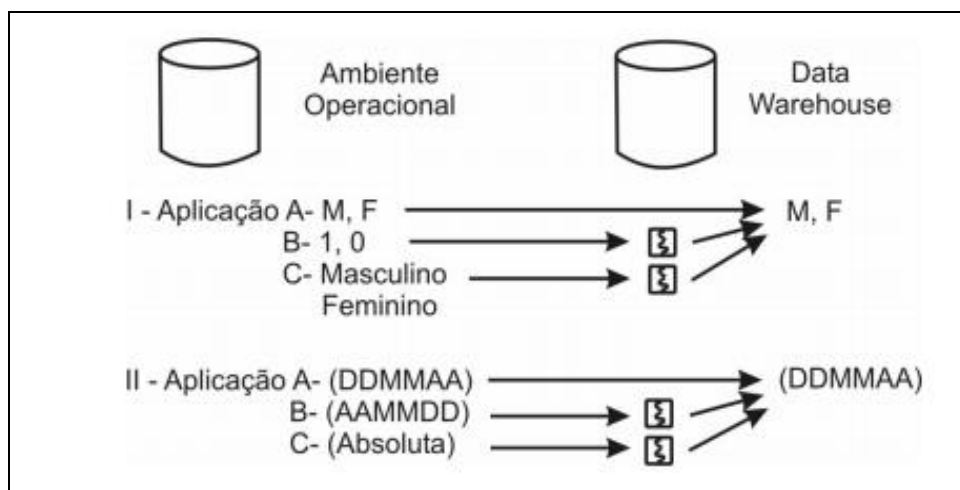


Figura 1 - Integração de dados
Fonte: Machado (2004, p. 23).

Pode-se acrescentar ainda outras características de extrema importância, como a não normalização dos dados que considera que esses podem ser redundantes; a abrangência que mantém uma série temporal, considerando que mais dados serão armazenados e dados resumidos, isto é, dados operacionais são agregados.

2.6 MODELAGEM DOS DADOS

Modelar é representar por meio de um modelo o ambiente observado, ou seja, através de diagramas e notações mostram-se como as informações estão relacionadas. A modelagem de dados é um modo eficiente de entender os dados. Baseando-se nas afirmações de Cougo (1997) entende-se que o propósito da modelagem é prover um registro apurado de alguns aspectos do mundo real para um contexto particular. É através da modelagem do banco de dados que o projetista pode eliminar redundâncias e fontes de informações inconsistentes. Não utilizar um modelo implica em um crescimento desorganizado, provendo altos custos e dificuldades no desenvolvimento de uma aplicação.

A modelagem deve sempre que possível levar em consideração a possibilidade de implementações futuras nas organizações. A capacidade de reconhecer antecipadamente estas necessidades dependerá em grande parte da experiência e conhecimento da equipe de desenvolvimento em relação às necessidades da organização. De acordo com a afirmação de Araújo (2011, p.01 *apud* Wen, 2007), a modelagem deve levar em consideração aspectos como a aquisição, arquivamento, padronização, comunicação e integração dos dados, além de armazenamento e recuperação das informações e análise dos dados para levantamento.

Segundo Hokama *et al.* (2004), afirma-se que modelos de dados podem ser classificados segundo a arquitetura que utilizam, sendo o relacional para sistemas transacionais e o modelo dimensional para sistemas analíticos. Para a construção de um DW deve-se definir inicialmente a modelagem dos dados que se enquadram, podendo ser Modelagem Entidade-Relacionamento ou Modelagem Multidimensional.

2.6.1 Modelagem Entidade-Relacionamento (ER)

O modelo de Entidade-Relacionamento (ER) serve de base para a construção do banco de dados, isto é, trata-se de desenhar abstratamente as relações que podem ser observadas no mundo real. Obtendo-se, assim, a visão planejada do que será implementado. Sabe-se que esse modelo utiliza três símbolos gráficos: entidades, relacionamentos e atributos, como exemplificado pela Figura 2.

- **Entidade**

Objeto que existe no mundo real, definido como: pessoa, lugar ou evento de interesse, classificadas por propriedades e características.

- **Relacionamentos**

Relacionamento é traduzido por linhas entre as entidades em um ER, que representa uma ação entre as mesmas, por meio de um verbo.

- **Atributos**

Descrevem as características de uma entidade como, por exemplo, nome, cor, tamanho, estado, altura.

- **Cardinalidade**

Indica quantas ocorrências de uma Entidade participam do relacionamento, sendo que o primeiro número representa a obrigatoriedade da ocorrência do elemento no relacionamento podendo ser 0 para não obrigatório e 1 para obrigatório, e o segundo número indica o número ocorrências da entidade, podendo ser 1 para exatamente uma ocorrência e n para indefinido número de elementos.

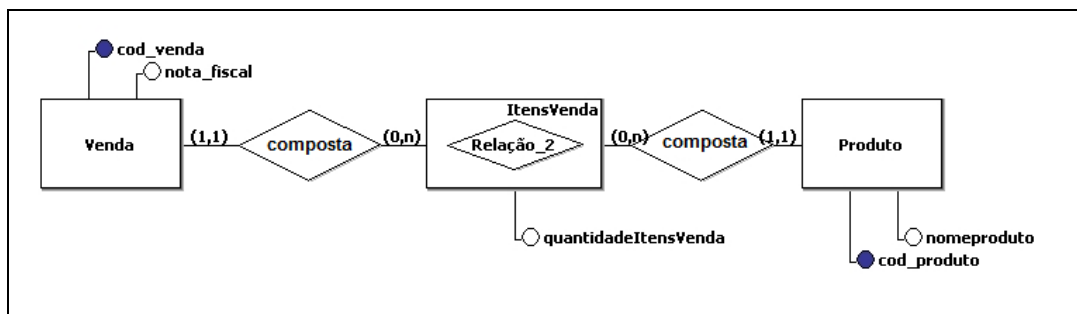


Figura 2 - Diagrama de Entidade e Relacionamento
Fonte: Autoria Própria.

Além disso, a entidade pode variar em supertipo e subtipo. O relacionamento entre entidade supertipo e entidade subtipo é utilizado quando uma entidade é a generalização de muitas entidades especializadas. Cada entidade subtipo herda os atributos da entidade supertipo, conforme explicado por Machado(2004). A Figura 3 esclarece o item.

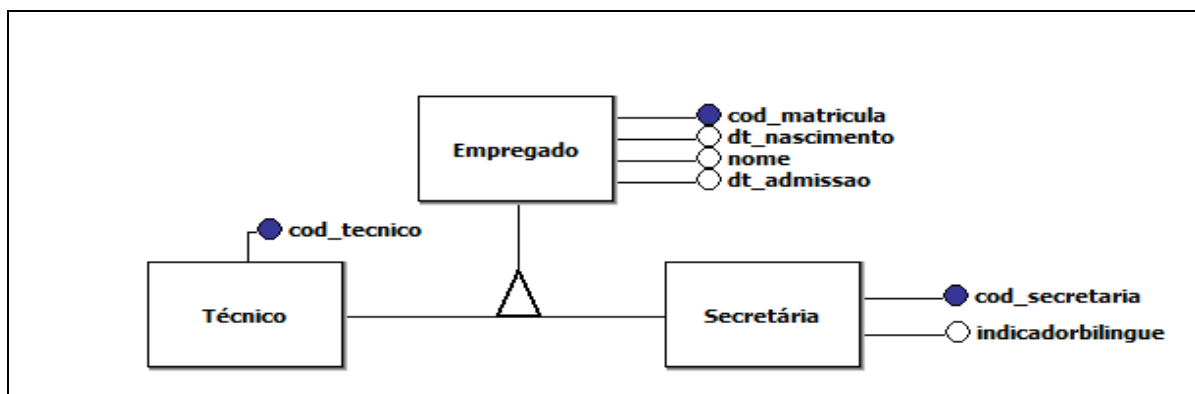


Figura 3 - Herança
Fonte: Autoria Própria.

Outra característica que se apresenta no modelo ER é a agregação, usada para definir uma nova classe a partir de um conjunto de classes que representam suas partes componentes. Ocorre em relacionamentos n-ários, como na Figura 4 abaixo.

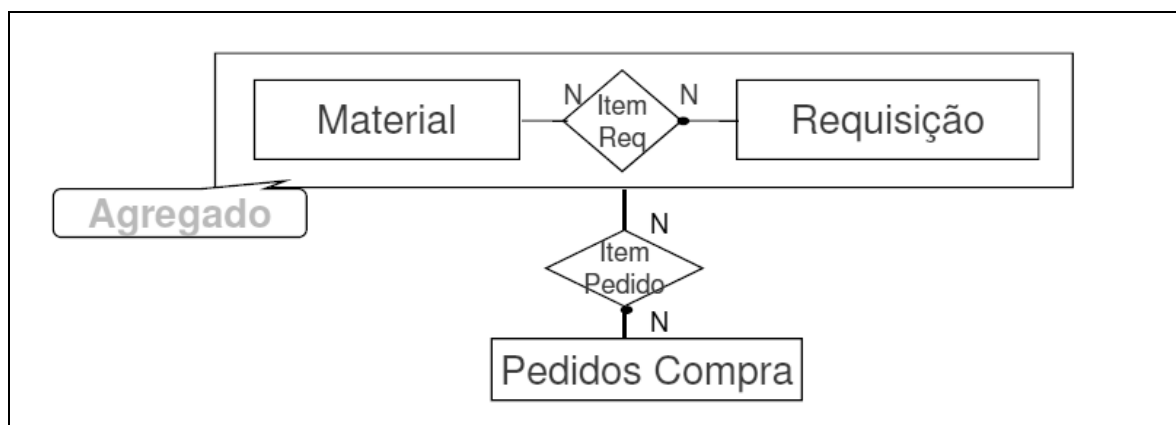


Figura 4 - Agregação
Fonte: Autoria Própria.

2.6.2 Modelagem Multidimensional

O modelo dimensional surgiu para atender sistemas de processamento analítico, com consultas para planejamento tático e estratégico da empresa. Atualmente a utilização desses sistemas pelo nível operacional das empresas vem crescendo, auxiliando o processo de tomada de decisões diárias. Normalmente, ele atende um pequeno número de usuários que realizam consultas planejadas e *ad-hoc*.

Para Machado (2004) modelagem multidimensional é uma técnica em que a concepção e a visualização descrevem aspectos comuns de negócios. Dividida em três elementos: fatos, dimensões e medidas.

- **Fatos**

Coleção de itens de dados composta de medidas. Implementado em tabelas denominadas tabelas de fato (*fact table*) e representado por um valor aditivo, valores numéricos denominados métricas ou medidas. Um fato é evolutivo e muda de acordo com o tempo, e possui três características, segundo Machado(2004): varia ao longo do tempo, possui valores numéricos de avaliação, seu histórico pode ser mantido e cresce com o passar do tempo.

Para denotar um fato é importante identificar indicadores matemáticos como participação, desempenho, evolução e índices. E ainda de acordo com Machado (2004), um fato não possui nenhum identificador percentual e não indica a inexistência do fato, porém é resultado de uma análise incompleta.

- **Dimensões**

Para se obter as dimensões dos fatos, tem-se quatro elementos participantes: quando foi realizada, onde foi realizada, quem realizou e o que foi. Obtendo essas respostas realiza-se a análise dimensional. As dimensões podem ser compostas por membros que podem conter hierarquias. Membros são as possíveis divisões ou classificações de uma dimensão. Por exemplo, a dimensão tempo, pode ser dividida nos seguintes membros: ano, trimestre e mês, e a dimensão localização em: cidade, estado e país;

- **Medidas**

São os atributos numéricos que representam um fato, ou seja, representam o desempenho de um indicador de negócios relativo às dimensões que participam desse fato. Uma medida é determinada pela combinação das dimensões que participam de um fato e estão localizados como atributos de um fato. Por exemplo, o valor em reais das vendas, o número vendido de unidades de produtos e a quantidade em estoque.

Kimball (1998) aponta um conjunto de nove pontos fundamentais no projeto da estrutura de um DW. São os chamados pontos de decisão, que constituem definições a serem feitas e correspondem, de fato, a etapas de um projeto:

- Os processos e, por consequência, a identidade das tabelas de fatos;
- A granularidade de cada tabela de fatos;
- As dimensões de cada tabela de fatos;
- Os fatos, incluindo os pré-calculados;
- Os atributos das dimensões;
- Como acompanhar mudanças graduais em dimensões;
- As agregações, dimensões heterogêneas, minidimensões e outras decisões de projeto físico;
- Duração histórica do banco de dados do *Data Warehouse*;
- A frequência com que se dá a extração e a carga para o *Data Warehouse*.

O esquema Estrela apresentado na Figura 5 é a estrutura básica de um modelo multidimensional, mas também se observa outros, como por exemplo, o floco de neve. Com a entidade central denominado fato (*fact table*) e o conjunto de entidades menores, as dimensões (*dimension table*).

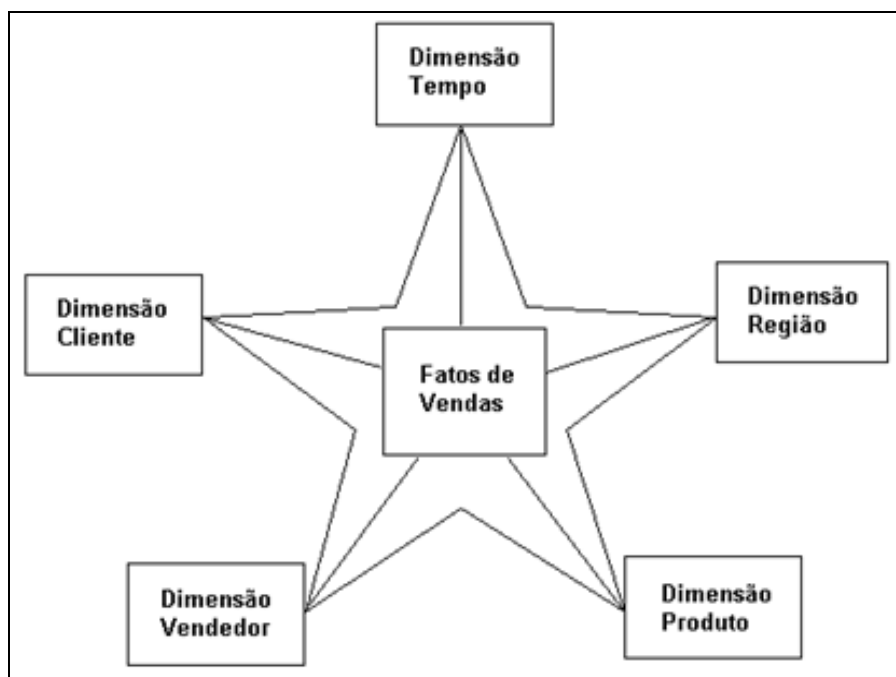


Figura 5 - Modelo Estrela
Fonte: Machado (2004, p. 93).

2.7 METADADOS

Metadados são considerados dados sobre dados, isto é, os dados sobre os sistemas que operam com estes dados. Um repositório de metadados é uma ferramenta essencial para o gerenciamento de um *Data Warehouse* no momento de converter dados em informações para o negócio. Um repositório de metadados bem construído deve conter informações como a origem dos dados, regras de transformação, nomes e apelidos e formatos de dados. Esse dicionário deve conter informações que adicionem valor aos dados.

Pode-se considerá-los como índice do conteúdo de um DW. Através dos metadados o usuário final compreende e analisa os dados, não sendo necessária uma busca baseada em tentativas (SANTOS, 2003).

Barquin (1997) classificou as vantagens dos metadados como a redução da possibilidade de erros, redução de complexidade, uma melhor captura do ambiente de dados

operacionais e fatoração do tempo. Na Figura 6 pode-se entender a ligação existente entre os usuários e os metadados.

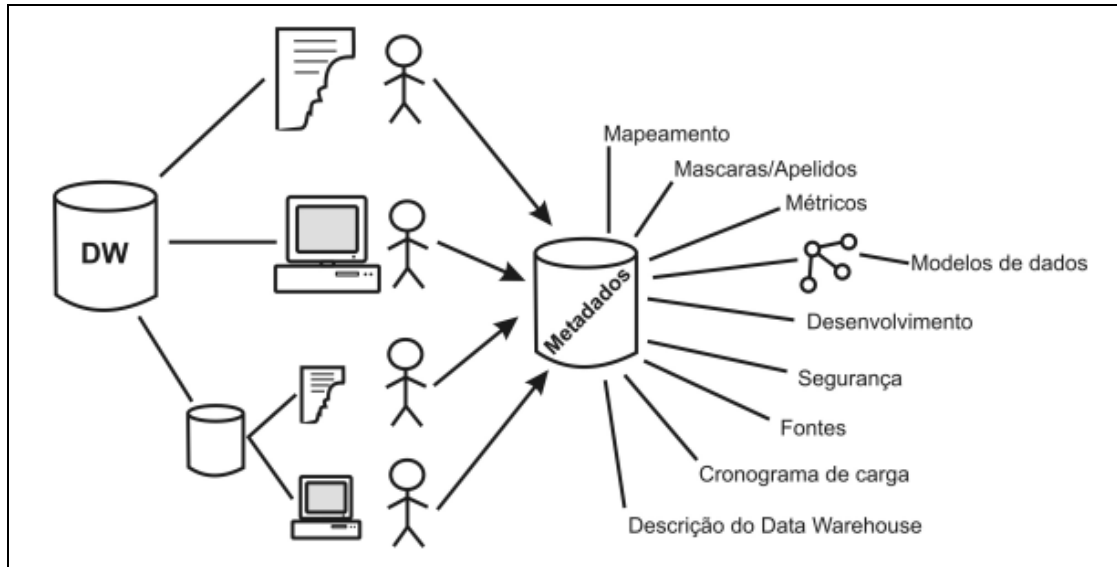


Figura 6 - Ligação dos Usuários aos metadados
Fonte: Machado (2004, p.27).

2.8 DATA MART

Data Mart (DM) é um subconjunto de DW, que contém dados para um setor específico da empresa. Um DW é composto por DMs. Conforme Machado (2004), a principal vantagem do seu emprego é o retorno rápido de informações, o que garante maior comprometimento do usuário final, com a possibilidade de avaliar os benefícios extraídos de seu investimento. É modelado em um esquema Estrela de acordo com as necessidades do usuário final.

2.9 GRANULARIDADE

Segundo Inmon (1999), a granularidade é o mais importante aspecto do projeto de um DW. Trata-se do nível de detalhe exigido dos dados. “Quanto mais detalhado os dados forem, mais baixo é o nível de granularidade” (GRAY, 1998, p. 01). O baixo nível de granularidade aumenta o volume de dados armazenados no DW e o alto nível limita as pesquisas, com poucos detalhes. É a granularidade que medirá o volume de dados do DW, com ela ocorre uma hierarquização de dados.

2.10 PROCESSO DE DATA WAREHOUSING

O processo de fazer *Data Warehouse* é denominado *Data Warehousing* e embora sua estrutura ainda esteja em evolução, conforme afirma Machado (2004) essa evolução surgiu da resposta da crescente complexidade desse ambiente e das dificuldades de integração entre os componentes, pode-se ter uma boa base com o esquema a seguir, na Figura 7.

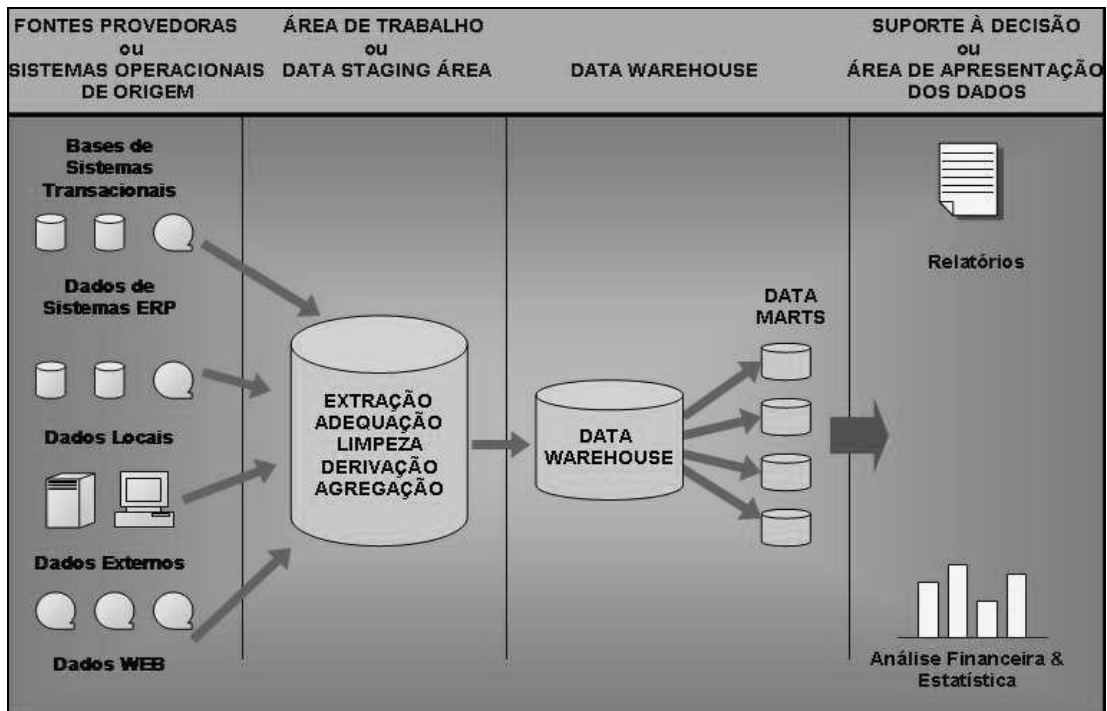


Figura 7 - Processo de DataWarehousing
Fonte: SunExpert Magazine (1998).

Sistemas Operacionais de Origem: são os sistemas operacionais de registro ou sistemas transacionais que capturam as transações da empresa. Também denominados Sistemas Legados ou OLTP. Onde é aplicado também o modelo utilizado no DW relacional ou multidimensional.

A data staging área: É tanto uma área de armazenamento como um conjunto de processos, e normalmente denomina-se ETL (*Extract – Transformation - Load*).

Data Warehouse e Data Mart: A área de apresentação dos dados é o local em que os dados ficam organizados, armazenados e tornam-se disponíveis para serem consultados diretamente pelos usuários, e o *Data Mart* trata somente de uma parte do *Data Warehouse*. Machado (2004, p. 44) afirma que “o *Data Warehouse* é um armazém de dados e os *Data Marts* são as prateleiras”.

Ferramenta de acesso a dados: O último componente principal do ambiente de *Data Warehouse* é a ferramenta de acesso a dados. O sistema OLAP está pronto para consultas.

2.11 TIPOS DE IMPLEMENTAÇÕES

Resumidamente, segundo Machado (2004), a implementação pode dividir-se em três: *Top Down*, *Bottom Up* e *Combine*.

A implementação *Top Down*, ou seja, de cima para baixo, requer um maior planejamento em que inicia com a extração, a transformação e a integração das informações dos sistemas. Logo após a extração dos dados, passa-se ao DW onde são extraídos os dados e metadados para os DMs.

A implementação *Bottom Up*, de baixo para cima, é popular e realiza-se a implementação dos DMs sem a realização do DW, que é construído de forma incremental. Geralmente é escolhida por executivos por ser de retorno rápido. Inicia-se com a extração, a transformação e a integração dos dados para um ou mais DMs independentes. Um grande problema é a dificuldade em garantir uma padronização onde ocorre a falha na elaboração incremental do DW, e ainda inconsistência nos DMs.

Por fim a implementação *Combine*, ou combinada, nada mais é do que a junção da *Top Down* com a *Bottom Up*. Uma modelagem de dados do DW de visão macro, sendo o passo seguinte à implementação de partes desse modelo, partes essas escolhidas por processos ou atividades da área de interesse dos DMs.

3 FERRAMENTAS DE DATA WAREHOUSE

O capítulo tem por objetivo apresentar as características de duas ferramentas candidatas a utilização no desenvolvimento do DW. A seção 3.1 traz a apresentação das ferramentas que serão estudadas. Na Seção 3.2 será explanada sobre a ferramenta *Microsoft® SQL Server 2008 Express*. Dando sequência, a Seção 3.3 evidenciará a ferramenta *Pentaho®* e finalmente a Seção 3.4 demonstrará a comparação e escolha da melhor ferramenta para este trabalho.

3.1 APRESENTAÇÃO

Segundo Inmon (1999), para gerenciar um DW é necessário possuir as seguintes habilidades:

- Projetista de Banco de Dados;
- Modelador de dados;
- Desenvolvedor;
- Político - compete pelos recursos necessários para construção do DW;
- Gerente de Sistemas;
- Planejador de Sistemas;
- Usuário final;

E para gerir todo esse sistema faz-se necessário a busca de ferramentas adequadas para aplicação de DW, isto é, um conjunto de tecnologias que ofereçam suporte ao ambiente de informação. Sabe-se que há no mercado atualmente, uma grande quantidade de ferramentas que tem por papel o *Business Intelligence* e que a falta de conhecimento de padronização estrutural e operacional dificulta a comparação.

A partir dessa informação e após um rápido estudo sobre as melhores ferramentas do mercado, foram escolhidas duas ferramentas para serem analisadas e, posteriormente, uma delas utilizada no desenvolvimento do estudo de caso. Serão apresentadas as ferramentas *Microsoft® SQL Server 2008 Express* e *Pentaho®*.

Além dos testes realizados com as ferramentas, a análise do melhor programa para ser utilizado no estudo de caso foi baseada nas doze (12) regras para avaliar uma ferramenta OLAP, apresentadas por Rodrigues (2010, p.05 *apud* Codd e Salley, 1993, p.12) – que podem ser observadas no quadro 3.

CARACTERÍSTICA	DESCRIÇÃO
Visão Conceitual Multidimensional	Suportar modelo de visualização multidimensional, possibilitar a manipulação dos dados por administradores de forma fácil e intuitiva.
Transparência	Ser transparente ao usuário e ter suporte a padrões abertos de documentos para que o usuário possa incorporar os dados onde ele precise.
Interoperabilidade	Fazer análises com base em vários tipos de SGBD, relacionais ou não relacionais, além de SGBDs legados.
Desempenho Consistente de Relatórios	Prover recursos para rever as estratégias de obtenção de dados para melhorar a performance.
Arquitetura Cliente/Servidor	Capacidade de funcionar em um ambiente cliente/servidor.
Manipulação dinâmica de dados esparsos	Ajustar de acordo com a distribuição dos dados de forma a oferecer a máxima eficiência no gerenciamento da memória.
Dimensionalidade genérica	Cada dimensão deve ser equivalente em estrutura básica e recursos adicionais.
Suporte multiusuário	Capacidade de vários usuários trabalharem ao mesmo tempo com a mesma análise.
Operações de dimensões cruzadas irrestritas	Capacidade de refazer todos os cálculos de fórmulas associadas sem que o usuário tenha que redefinir explicitamente os cálculos.
Manipulação intuitiva de dados	A navegação entre as várias dimensões deve ser feita através de ações diretas sobre os dados da análise, sem exigir o uso de menus ou múltiplas interações do usuário.
Relatórios flexíveis	A disposição das colunas e a ordenação das linhas do relatório devem ser livres de forma a facilitar a comparação dos dados.
Níveis de dimensões e agregações ilimitados	Gravar pelo menos 15 dimensões e preferivelmente 20 dimensões.

Quadro 3 - Doze Regras do Ambiente OLAP
Fonte: Adaptado de Rodrigues (2010).

3.2 MICROSOFT® SQL SERVER 2008 EXPRESS

O *Microsoft® SQL Server 2008 Express* é uma edição gratuita do *SQL Server*, ideal para o aprendizado, o desenvolvimento e a potencialização de aplicações pequenas de estação de trabalho, *web* e servidor.

A ferramenta possui suporte a *store procedure*, *trigger*, *function* e *view* e armazena todos os tipos de dados com suporte nativo a dados relacionais, *XML*, *FILESTREAM* e espaciais. Além de ser vantajoso por apresentar um grande desempenho, usabilidade, visualização e integração ao *Microsoft 2007 Office System* no *SQL Server Reporting Services*. Possibilita a simplificação do desenvolvimento, potencializando as habilidades existentes em *T-SQL*, *ADO.NET Entity Framework* e *LINQ* e é integrado ao *Visual Studio* e ao *Visual Web Developer*.

3.2.1 Componentes

Segundo o manual³ disponível no sítio para *download* da ferramenta, a *Microsoft® SQL Server 2008 Express* é composta por diversos recursos que garantem a complexidade e qualidade do *software*, as quais são apresentadas a seguir:

O Mecanismo de Banco de Dados é o serviço principal para armazenamento, processamento e segurança de dados. Ele fornece acesso controlado e processamento rápido de transações para atender aos requisitos dos aplicativos de consumo de dados mais exigentes dentro da empresa.

O *Analysis Services* suporta *OLAP*, permitindo projetar, criar e gerenciar estruturas multidimensionais que contenham dados agregados de outras fontes de dados, como bancos de dados relacionais. Também permite projetar, criar, e visualizar modelos de mineração de dados. Estes modelos de mineração podem ser construídos a partir de outras fontes de dados utilizando uma diversidade de algoritmos de mineração de dados padrão da indústria.

O *Integration Services* é uma plataforma para construir soluções para integração de dados de alto desempenho, inclui também pacotes que fornecem processamento de extração, transformação, e carregamento (ETL) para armazenamento de dados.

O *Master Data Services* é a origem de dados mestres para a organização. Ao integrar sistemas analíticos e operacionais distintos com o *Master Data Services*, garante-se que todos os aplicativos da organização dependam de uma fonte de informações central e precisa. Usando o *Master Data Services*, cria-se uma única fonte de dados mestres e mantém um registro auditável desses dados conforme eles forem alterados com o tempo.

A replicação é um conjunto de tecnologias para copiar e distribuir dados e objetos de um banco de dados para outro e, em seguida, sincronizar entre os bancos de dados para manter a consistência. Ao usar a replicação, é possível distribuir dados para diferentes locais e para usuários remotos e móveis através de redes locais e de longa distância, conexões discadas, conexões sem-fio e a Internet.

O *Reporting Services* fornece às corporações a funcionalidade de relatórios *online* possibilitando criar relatórios que se conectam a conteúdos de várias fontes de dados, permite também, publicar os relatórios em diversos formatos e, além disso, centralizar o gerenciamento de segurança e de assinaturas.

³ Manual disponível no sítio eletrônico < <http://msdn.microsoft.com/pt-br/library/bb522607.aspx> > .

O *SQL Server 2008 R2* oferece novo recurso de *Business Intelligence* de autoserviço através da integração com outros produtos e tecnologias.

O *Service Broker* auxilia os desenvolvedores a construir aplicativos de banco de dados evolutivos e seguros. Esta nova tecnologia do Mecanismo de Banco de Dados fornece uma plataforma de comunicação baseada em mensagens que permite aos componentes de aplicativos independentes funcionarem como um todo. O *Service Broker* inclui infraestrutura para programação assíncrona que pode ser usada em apenas um banco de dados ou apenas uma instância, e também para aplicativos distribuídos.

3.2.2 Caso de Sucesso

Empresa

Trip Linhas Aéreas®

Situação

A Trip Linhas Aéreas® cresceu muito rapidamente e com o crescimento veio também a necessidade de mudar o fornecedor de aviação para um que oferecesse solução de aviação internacional além da nacional já existente.

Solução

Uma das soluções foi a utilização do *SQL Server 2008 R2* em que a persistência dos dados é viabilizada pelo *Entity Framework*, que proporcionou um grande ganho de produtividade.

Todas as funcionalidades dos sistemas possuem testes automatizados de unidade e funcionais, criados através do *Visual Studio 2010* e executados nas compilações diárias automatizadas. Foram desenvolvidos relatórios adicionais com o *Reporting Services* que possibilitou uma visão mais personalizada das informações do processo, junto ao *Team Foundation*. O *Team Explorer* foi utilizado em conjunto com a versão 2010 de *Excel* e *Project*, para facilitar a visualização e planejamento dos itens de trabalho cadastrados. O padrão de codificação e nível de qualidade são garantidos pela utilização do *Code Analysis* e *StyleCop* de forma integrada aos *check-ins*.

Benefícios

Hoje a Trip Linhas Aéreas® em conjunto com o novo fornecedor de aviação internacional e as novas soluções do portal, poderá aumentar a oferta de serviços para o usuário final e está em igualdade competitiva com as maiores companhias de aviação mundiais e, além disso, já chega com diferenciais em termos de facilidades de uso e

flexibilidade de escolha de vôos e de poltronas no novo portal, que ficou ainda mais fácil de usar e com ótimo desempenho, reduzindo o tempo necessário para realizar as operações de reservas ou compra de passagens.

3.3 PENTAHO®

Uma ferramenta fundada em 2004, que se baseia em tecnologia *Java*, sendo de código aberto e gratuito. Com ela pode-se realizar consultas, relatórios, análise de informações (OLAP), ETL e mineração de dados.

3.3.1 Componentes

Baseado no manual⁴ de utilização da aplicação disponível no sítio eletrônico da empresa, define-se os componentes da ferramenta como:

Pentaho Data Integration conhecido com *Kettle* é uma ferramenta de código aberto para extração, transformação e carga (ETL) de dados. *Pentaho Analysis Services* também conhecido como *Mondrian OLAP Server* é uma ferramenta de código aberto para *OLAP*.

Pentaho Reporting gerador de relatórios, derivado do projeto *JFreeReport*.

Pentaho DataMining derivado do projeto *WEKA*, um conjunto relacionado a ferramentas de mineração de dados.

Pentaho Dashboard são coleções de gráficos e/ou relatórios geralmente agrupados com um mesmo tema para facilitar a visualização e compreensão das informações.

Pentaho for Apache também conhecido como *Pentaho BI Suite*, é a versão para *Hadoop*, uma plataforma de *software* em *Java* de computação distribuída voltada para *clusters* e processamento de grandes massas de dados.

3.3.2 Caso de Sucesso

Empresa:

Dia % Brasil®

Visão geral de implantação:

A empresa buscava o desenvolvimento gratuito de um portal comercial de *Business Intelligence*. Além disso, necessitava de um DW, para obter informações da matriz na

⁴ Manual disponível o sítio eletrônico <<http://www.pentaho.com.br>>.

Espanha e outros sistemas. Finalmente, requeria organização, catalogação e compartilhamento de informações estratégicas através de um sistema *web*, com monitoramento das vendas da empresa.

Solução Pentaho:

Suite Business Intelligence Pentaho®.

Resultados:

Decisões tomadas segundo relatórios e informações geradas através da ferramenta, além do ganho com a rapidez da geração dos dados, relatórios e informações.

3.4 COMPARAÇÃO

Baseada nas pesquisas de seleção das ferramentas, que mostrou suas funcionalidades, componentes e recursos adicionais, encontrou-se um leque para comparações e escolha da melhor ferramenta.

A arquitetura da suíte da *Pentaho®* é voltada pra *web* e construída em *Java*, sendo isso um de seus principais diferenciais. Já a *Microsoft® SQL Server 2008 Express* é ideal para *desktop*.

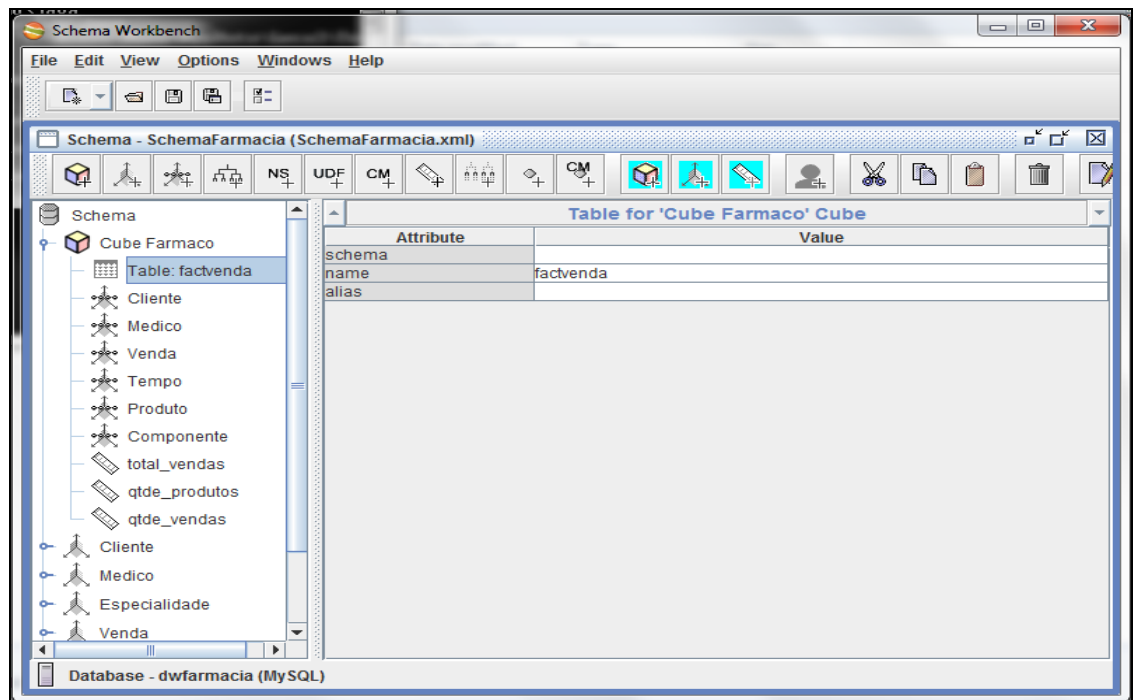


Figura 8 - Pentaho Schema Workbench
Fonte: Adaptado de Workbench.

Com a *Pentaho*®, a geração de cubos é através do mapeamento do arquivo XML e a ferramenta utilizada para essa etapa é a *Pentaho Schema Workbench* (PSW), como pode ser visualizado na Figura 8 e a ferramenta utilizada para a análise do cubo é a *Analysis View* ou *Mondrian*. A suíte *Microsoft*® não utiliza o mapeamento do arquivo XML. O processo de análise e geração do cubo foi realizado usando a ferramenta *SQL Server 2008 Analysis Services*.

O *Pentaho* executa em um servidor compatível com J2EE. A ferramenta utiliza XPDL para definição dos processos em XML e os repositórios podem ser em *MySQL*, *FireBird*, *Oracle*, *Postgre* e outros. A *Microsoft*® *SQL Server 2008 Express* requer o uso do repositório *SQL Server* somente.

A ferramenta *Pentaho Data Integration* (PDI) ou *Kettle* utiliza-se no processo de ETL. O DM produzido foi armazenado no SGBD livre *MySQL*. Com a suíte *Microsoft*, o processo de ETL é realizado usando a ferramenta *SQL Server 2008 Integration Services* e para o armazenamento do DM é utilizado o SGBD *Microsoft SQL Server 2008*.

O quadro 4 apresenta as principais diferenças encontradas entre as ferramentas.

CARACTERÍSTICA	PENTAHO	SQL SERVER 2008
Multi-Plataforma	Sim	Não
Disponível na Internet	Sim	Não
Adaptação funcional de componentes	Sim	Não
Exportação formato livre (ODT)	Sim	Não
Facilidade de Uso	Menor	Maior
Interface Personalizável	Maior	Menor
Suporte técnico/Documentação	Menor	Maior
Definição de níveis/perfis	Sim	Não
Integração com outras ferramentas	Maior	Menor

Quadro 4 - Diferenças entre *Pentaho* e *SQL Server 2008*

Fonte: Autoria Própria.

A *Microsoft*® *SQL Server 2008 Express* tem um maior suporte e conta com atualizações frequentes, diferente do *Pentaho* que no lugar do suporte, possui fóruns e comunidade da ferramenta.

Além disso, a *Microsoft*® *SQL Server 2008 Express* por estar estabilizada no mercado e anos de atuação em empresas, tem maior credibilidade do que outras ferramentas novas, como a *Pentaho*. No entanto, para o estudo de caso escolhido, será utilizada a ferramenta *Pentaho*, tanto pelas vantagens já indicadas como pela valorização de uma ferramenta nova no mercado e *open source*.

4 ESTUDO DE CASO

O processo de DW foi aplicado em uma base de dados de uma farmácia de manipulação e está descrito neste capítulo. O capítulo apresenta o trabalho desenvolvido com a ferramenta escolhida para o processo de DW. A Seção 4.1 apresenta a descrição da empresa e os objetivos esperados com a aplicação do DW. A Seção 4.2 explana sobre o pré-processamento, ou seja, a identificação do conhecimento que deverá ser devolvido com a finalização do DW. A Seção 4.3 explica o modelo lógico, isto é, o modelo multidimensional. Na Seção 4.4 são apresentadas as ferramentas que a suíte *Pentaho* utiliza e suas configurações e a Seção 4.5 finaliza com a implementação propriamente dita na suíte *Pentaho*.

4.1 DESCRIÇÃO

A IONSFARMA farmácia de manipulação e homeopatia, localizada em Ponta Grossa – Paraná desde 1999 é responsável por vendas de produtos comerciais e manipulados. Medicamentos manipulados são preparados em uma farmácia magistral autorizada pela Vigilância Sanitária e contém o princípio ativo e a dose prescrita pelo médico. Estes medicamentos são apresentados pelas farmácias de manipulação como alternativas às doses padrões disponibilizadas pela indústria, ou seja, a produção de medicamentos é personalizada em relação à dose, como também a combinação de princípios ativos em uma mesma cápsula ou outras formas de apresentação do medicamento (ANVISA, 2007).

Este é um segmento de várias particularidades e desafios constantes, muitas vezes impostos pelo órgão regulamentador ANVISA (Agência Nacional de Vigilância Sanitária) como:

- Controle de Preço;
- Divulgação dos produtos na mídia;
- Promoção médica limitada a regras de conduta;
- Normas de exposição dos produtos nas farmácias;
- Dificuldade de obtenção de registros de novos medicamentos.

Essa exigência cada vez mais acentuada das autoridades sanitárias evidencia a preocupação com a qualidade dos produtos manipulados. Assim, para se manter competitiva no mercado, as empresas desse ramo de negócios necessitam de informações que possibilitem maior conhecimento de seu negócio, auxiliando na tomada de decisão na empresa.

O objetivo geral é combinar as necessidades de informação da comunidade de usuários com os dados que realmente estão disponíveis, compondo o repositório de dados.

4.2 PRÉ-PROCESSO

Inicialmente é feito o processo extração de subsídios frente a frente com o usuário, uma atividade indispensável para recolhimento de informações relevantes que a organização queira obter. Uma maneira para realizar esse processo é através de *brainstorming* ou em entrevistas com o cliente.

Neste trabalho, a coleta de dados foi realizada junto à proprietária e administradora da empresa, requerendo informações vitais para gerar o DW. Foram identificadas as seguintes questões que o DW deverá responder:

- Qual o faturamento por médico, em um mês ou um ano?
- Qual o faturamento por especialidade médica?
- Qual o faturamento de um tipo de componente?
- Qual o faturamento por cliente em um mês ou um ano?
- Qual o componente mais vendido por especialidade médica?
- Qual o mês de maior e menor faturamento em 2011?
- Qual o faturamento em 2010?

Após definidos os questionamentos, é realizada a análise para dar início ao processo de modelagem. A Figura 9 mostra o ciclo de vida do DW em que inicialmente é realizado o planejamento do projeto, definindo as necessidades do negócio. Segue-se com os processos sequenciais mostrados no diagrama, sendo esses o projeto técnico de arquitetura, a seleção e instalação da ferramenta, a modelagem multidimensional, o projeto físico, a preparação de dados, a aplicação analítica e o desenvolvimento da aplicação. Após o término dessas fases, é feita a implementação e a manutenção do DW.

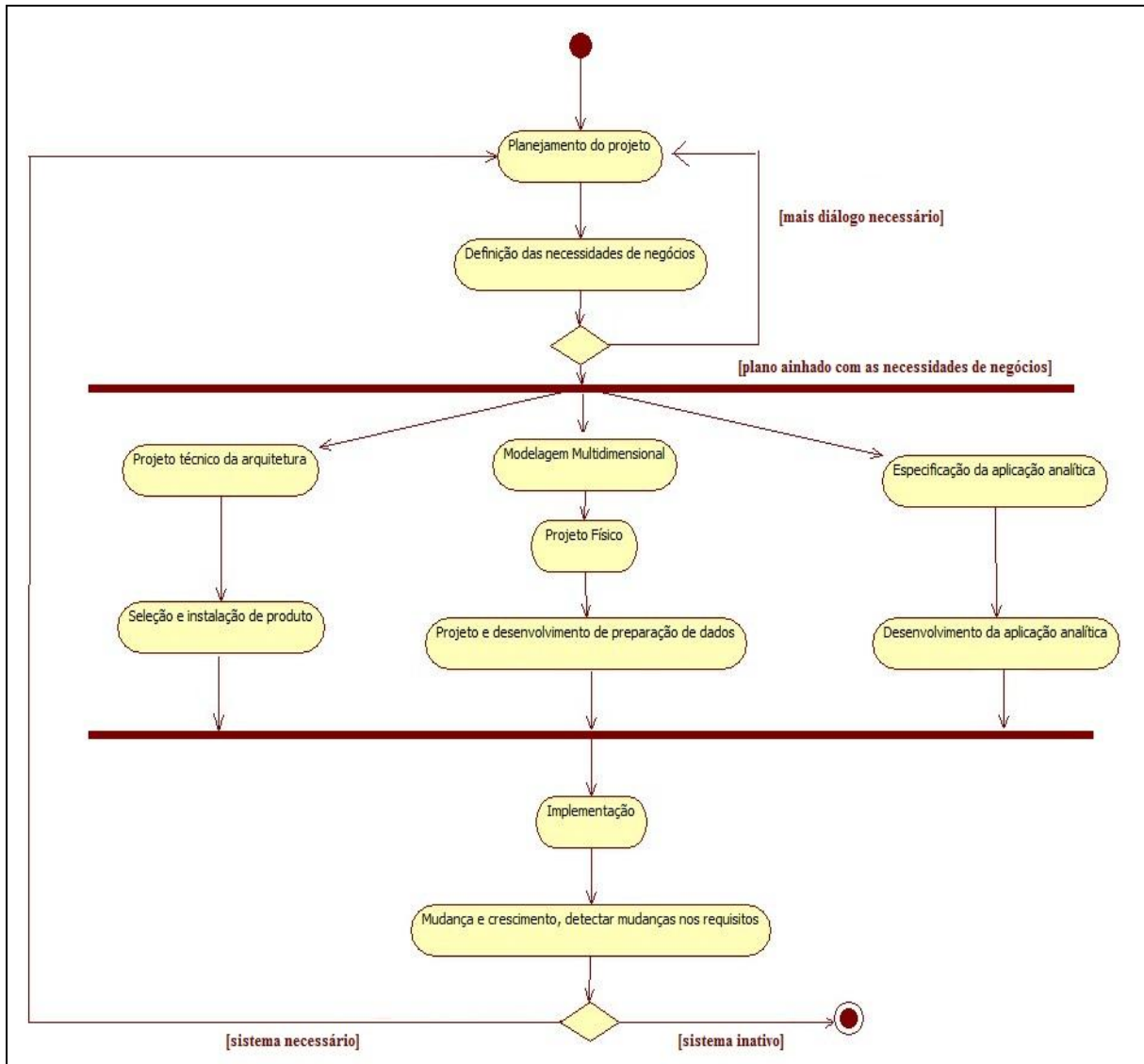


Figura 9 - Ciclo de vida do *Data Warehouse*
 Fonte: Teorey, *et al.* (p. 157, 2006).

4.3 PROJETO LÓGICO

Para dar início ao projeto lógico foi realizada uma análise minuciosa do Modelo de Entidade e Relacionamento do banco da empresa, conforme Anexo A, juntamente com as questões citadas na seção 4.1 para definir assim, o Modelo Multidimensional ilustrado na Figura 10. Foram definidas as seguintes dimensões: Cliente, Médico, Especialidade, Componente, Produto, Venda e Tempo. E o fato foi definido como: Venda.

Foi escolhido o processo de modelagem departamental, ou seja, um *Data Mart*, focando assim em um único problema da empresa. Diante das pesquisas, concluiu-se que o departamento propício seria vendas, com foco no faturamento.

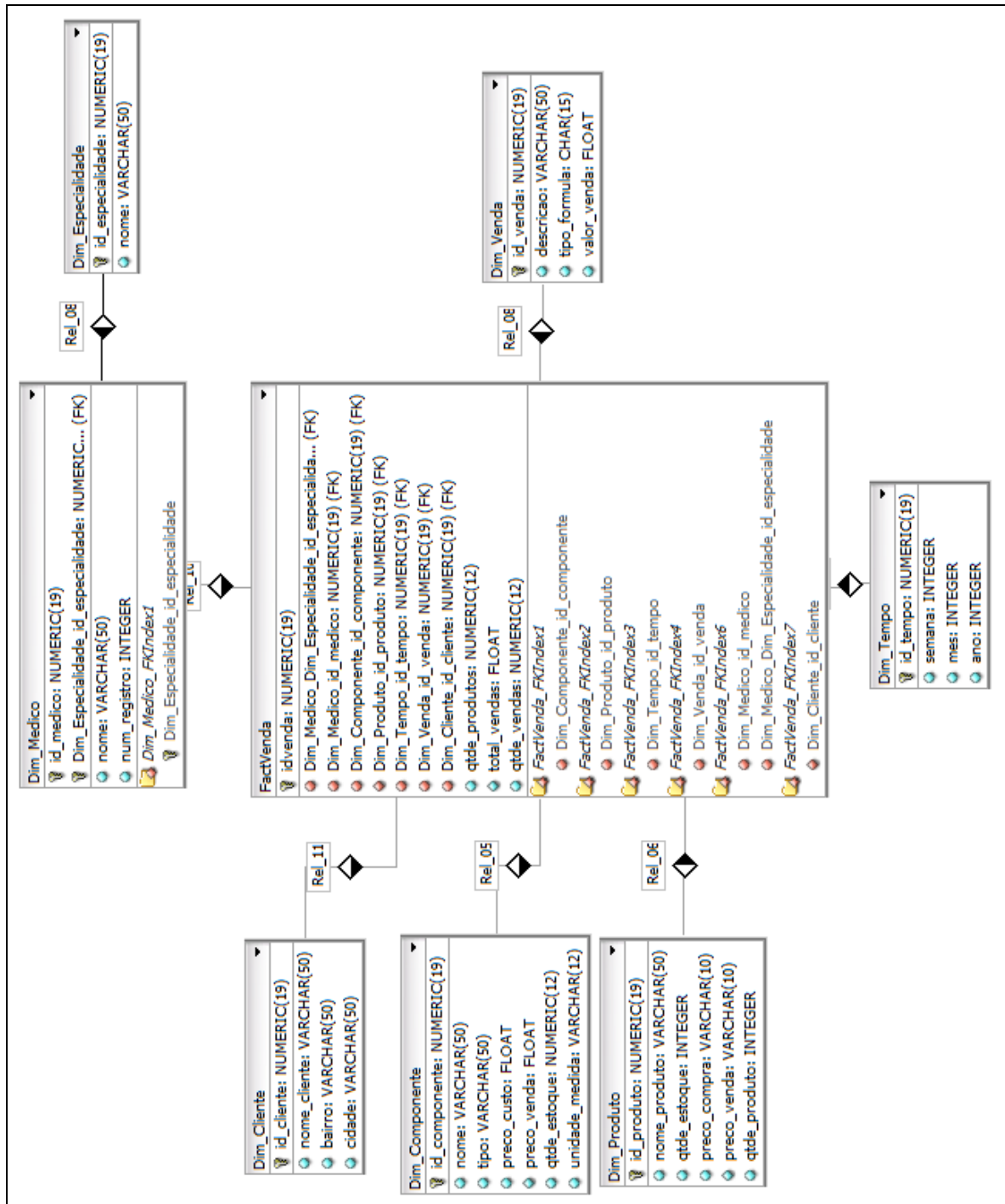


Figura 10 - Modelagem Multidimensional.

Fonte: Autoria Própria.

4.4 FERRAMENTAS E SUAS CONFIGURAÇÕES

Para a implementação do DW é necessário baixar a ferramenta *Pentaho*, que é composta por vários pacotes. Em cada pacote o seu executável é um arquivo *.bat* e para inicializar qualquer processo deve-se inicializar o servidor; um deles localizado dentro do

pacote *biserver-ce* (*start-pentaho.bat*) e o outro é localizado no pacote *administration-console* (*start-pac.bat*).

No estudo de caso foram usados os seguintes pacotes, que podem ser encontrados facilmente para *download* em < <http://sourceforge.net/projects/pentaho/files/> > acessado em 21 de novembro de 2011:

- ***biserve-ce***: para acessar a ferramenta, deve-se digitar no navegador o seguinte endereço: *http://localhost:8080/pentaho* e como usuário usa-se “Joe” e senha “password”.
- ***administration-console***: para acessar a área administrativa, deve-se digitar no navegador o seguinte endereço: *http://localhost:8099*. Usuário “admin” e senha “password”.
- ***schema-workbench***: para iniciar a ferramenta, clica-se no arquivo *workbench.bat*.
- ***data-integration***: para iniciar a ferramenta, utiliza-se o arquivo *spoon.bat*.

4.5 DESENVOLVIMENTO DO ESTUDO DE CASO

Para facilitar o entendimento do processo, desenvolveu-se um diagrama de atividades, auxiliando assim, na visualização dos passos a realizar por meio da utilização da tecnologia DW, ilustrado na Figura 11.

4.5.1 Atividade 1 - Criação da base de dados

Com o auxílio da ferramenta *SQLDbx* – um IDE para desenvolvimento SQL focado em ambientes de servidores heterogêneos – e utilizando o *driver* do banco de dados *MySQL*, criou-se o repositório de dados com as tabelas mostradas na Figura 10.

4.5.2 Atividade 2 - Gerar XML do modelo multidimensional

Nessa etapa realiza-se a conexão da base de dados nas ferramentas de administração: *administration-console* e *biserver-ce*. O fato é que, dependendo do banco de dados (BD) a ser utilizado mudam-se as configurações.

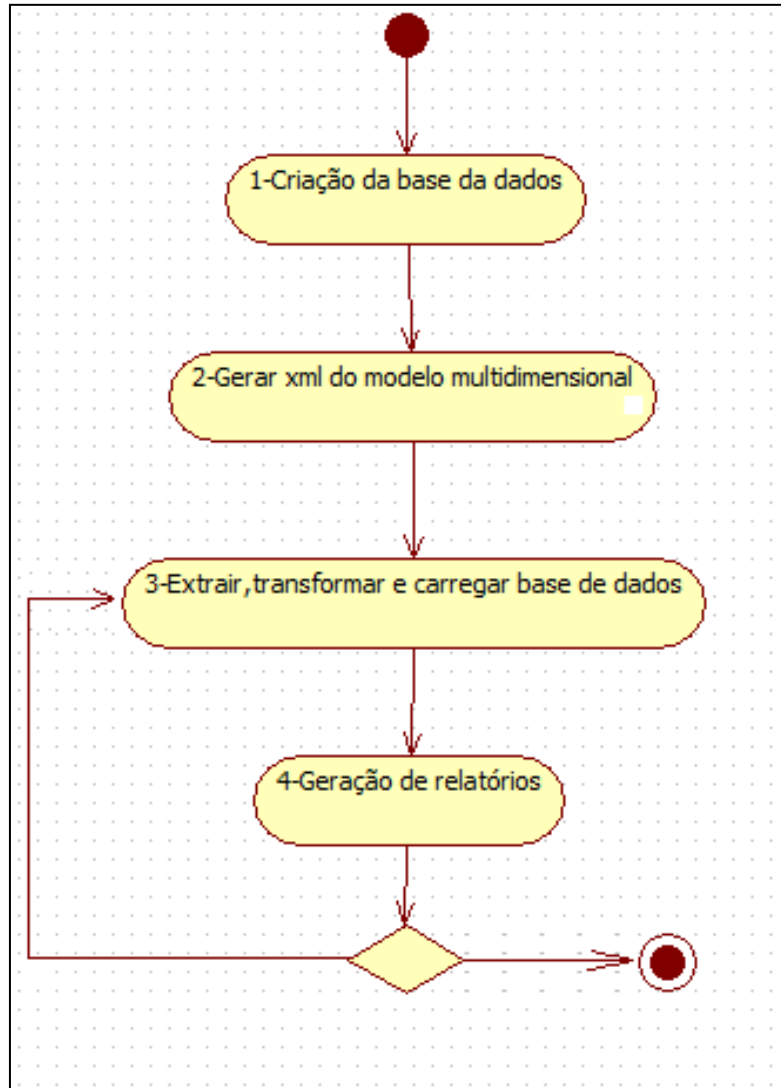


Figura 11 - Diagrama do processo de DW
Fonte: Autoria Própria.

A Figura 12 demonstra como ocorre a configuração do BD na ferramenta *administration-console*. No campo *driver-class* seleciona o *driver* que será utilizado. No campo *user-name* digita-se o nome do usuário e no campo *password* a senha do usuário. No campo *url* coloca-se o caminho do banco de dados.

Para conectar o BD no *biserver-ce*, realiza-se um processo semelhante, como indicado na Figura 13 completando os campos *host name* com “*localhost*”, *database name*: nome do banco de dados, *port number* é o número da porta que o banco de dados usa, além do usuário e da senha.

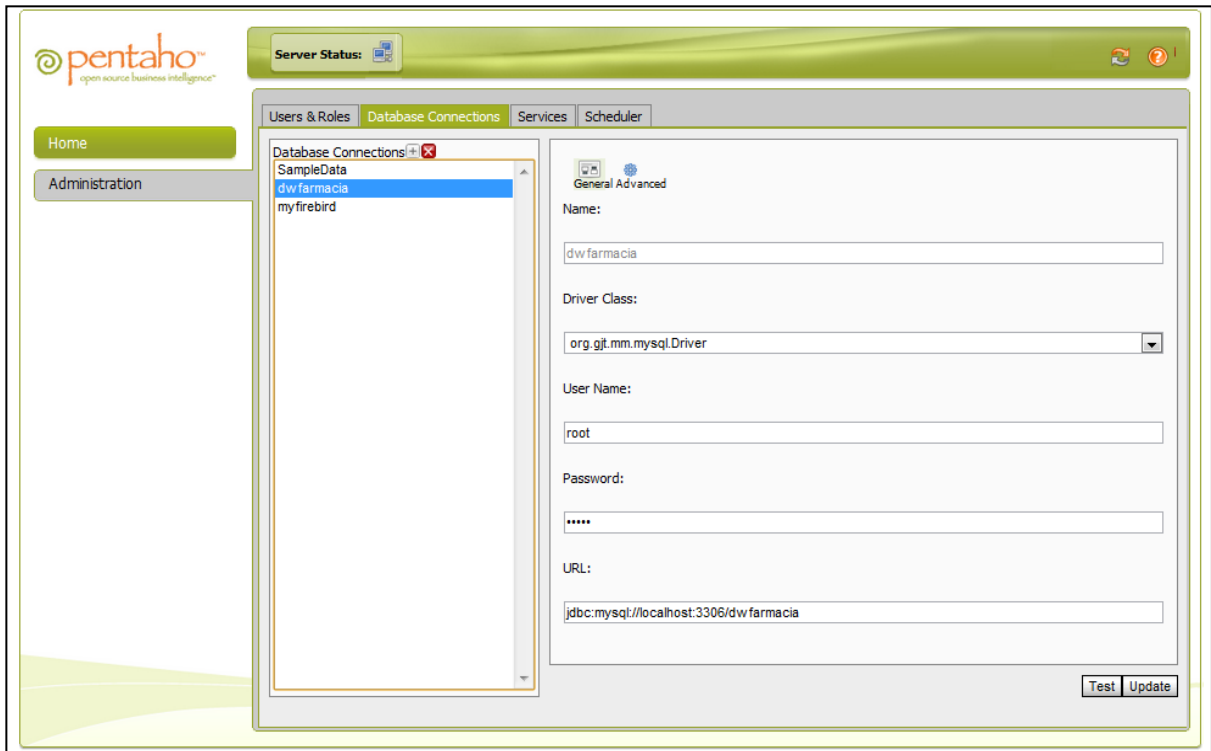


Figura 12 - Configuração do BD
Fonte: Autoria Própria.

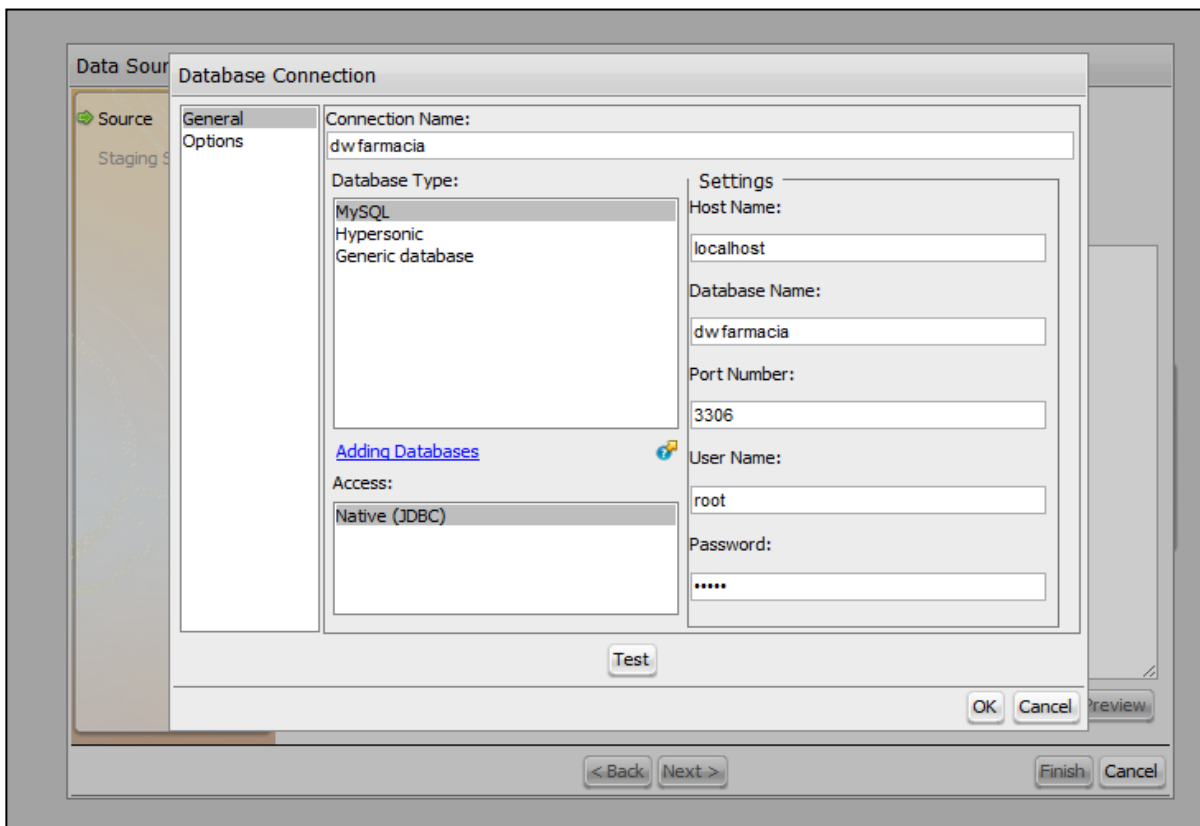


Figura 13 - Conexão com o banco de dados
Fonte: Autoria Própria.

Pode-se agora gerar o XML da modelagem multidimensional, para isso é usada a ferramenta *Schema Workbench*, onde é feito os seguintes processos: conexão com o BD, transformação do modelo multidimensional em XML e publicação na *web*. Primeiramente deve-se criar a conexão com o banco de dados completando os campos representados na Figura 14.

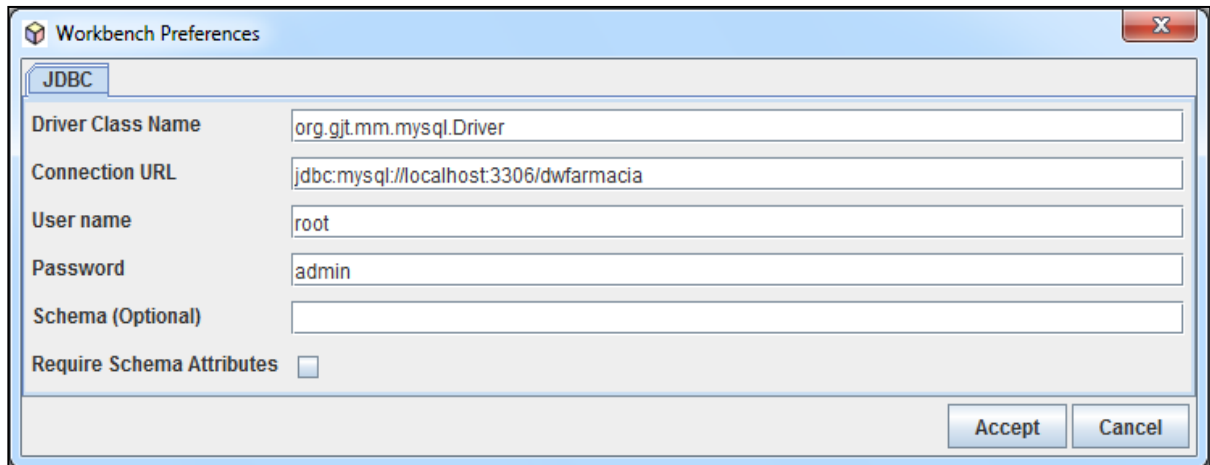


Figura 14 - Conexão com o BD
Fonte: Adaptado de *Workbench*.

Na Figura 15 é a representação da criação do modelo multidimensional, onde cria-se as tabelas, juntamente com seus atributos diferenciando fatos de dimensões. A Figura 16 mostra o XML pronto e a Figura 17 demonstra como publicar o XML, sendo necessário completar os campos *Pentaho* ou *JNDI Data Source*.

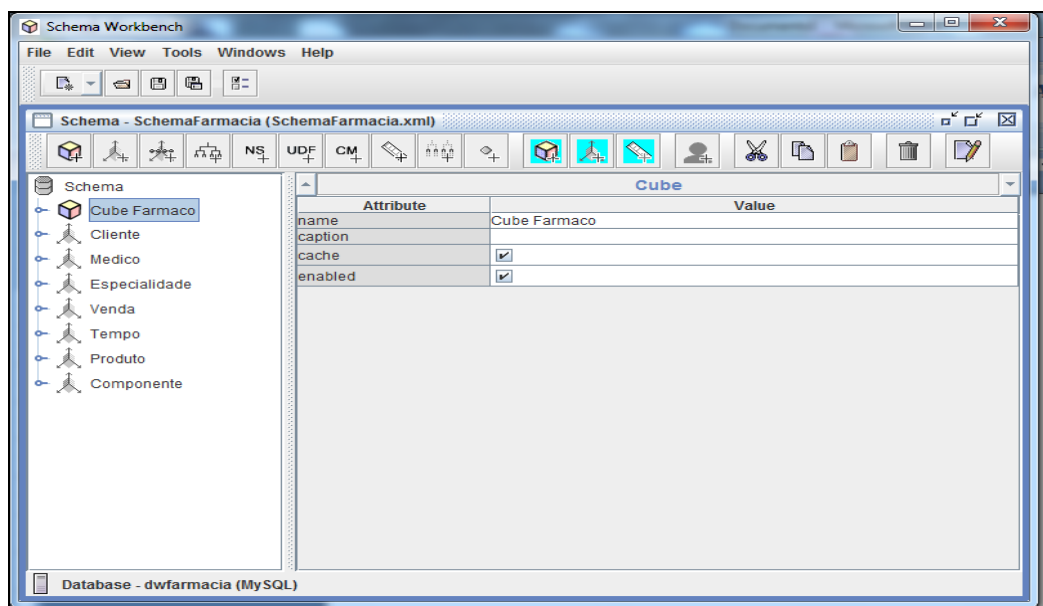


Figura 15 - Modelo Schema
Fonte: Adaptado de *Workbench*.

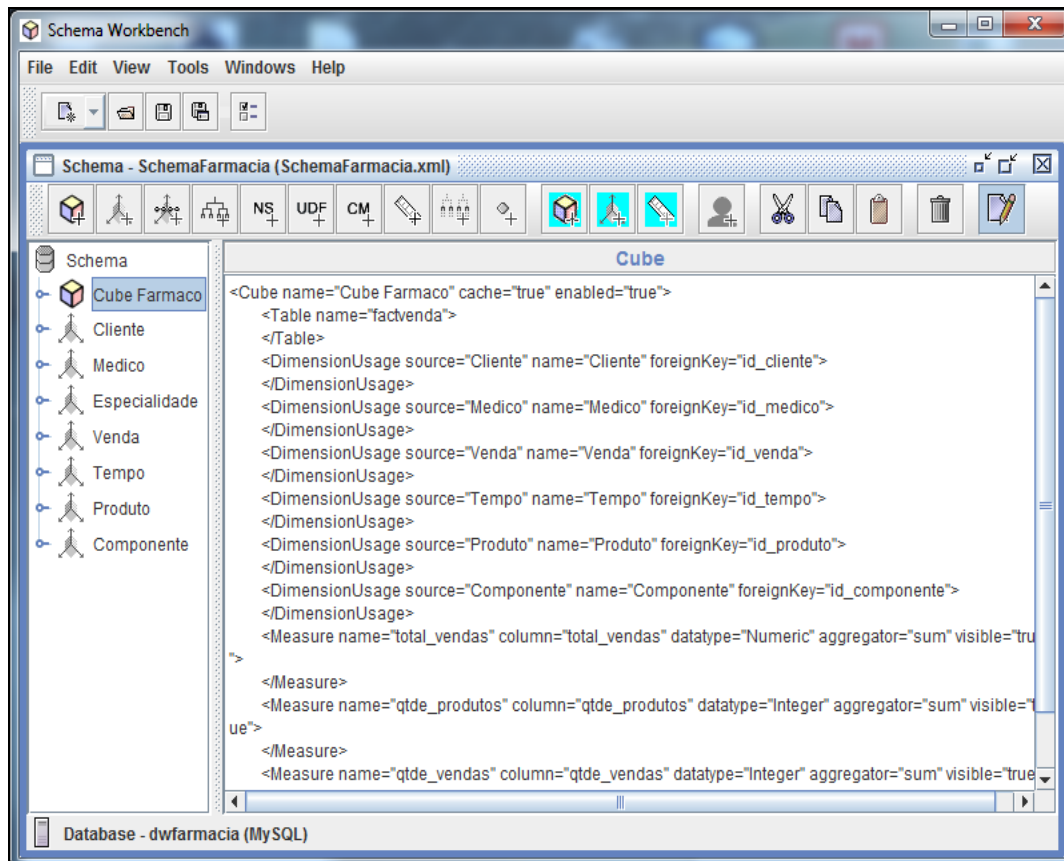


Figura 16 - XML do Cubo
Fonte: Adaptado de Workbench.

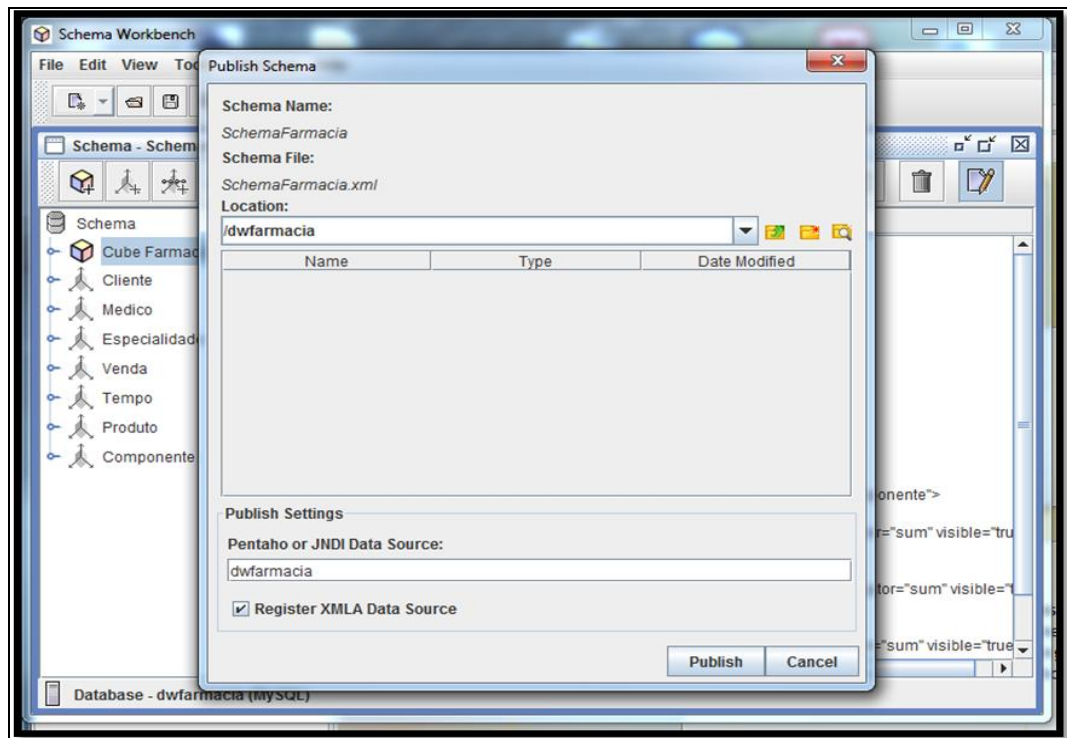


Figura 17 - Publicação do XML
Fonte: Adaptado de Workbench.

4.5.3 Atividade 3 - Extrair, transformar e carregar a base de dados

Para essa fase utiliza-se o pacote *data-integration*, em que é realizado o processo de ETL (*extract, transform, load*), ou seja, a extração, a transformação e o carregamento dos dados no repositório. De início, é necessário conectar o repositório na aplicação, o que é demonstrado na Figura 18.

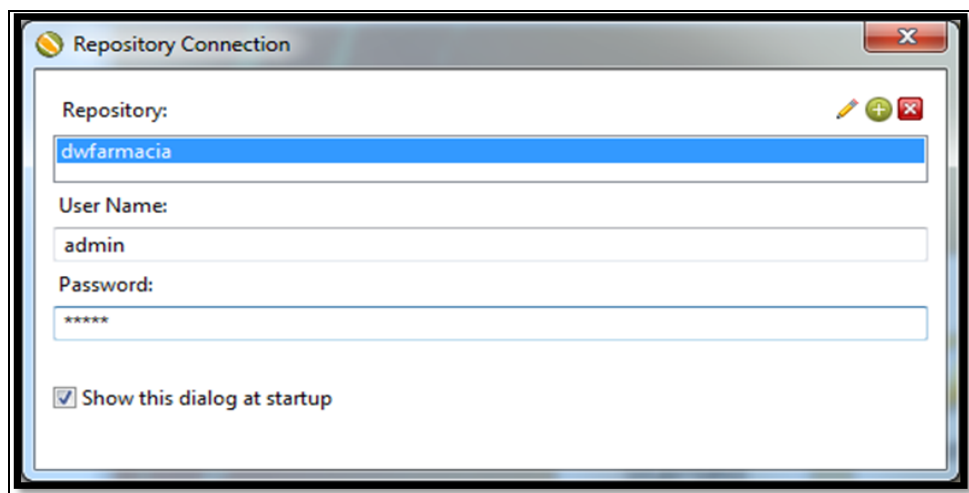


Figura 18 - Conexão do banco na Data Integration
Fonte: Adaptado da Spoon.

Após isso, deve-se iniciar a transformação dos dados. No estudo de caso, o desenvolvimento do DW foi realizado por partes: algumas dimensões do modelo multidimensional passaram por um processo de transformação; em outros momentos, mais de uma dimensão passaram pelo mesmo processo – como no caso da dimensão médico e dimensão especialidade – apresentadas na Figura 21.

Os processos de transformação, na maioria dos casos, consistem na ordenação dos dados e logo em seguida, na realização da carga dos mesmos no repositório. Adicionalmente em alguns processos de transformação foram feitos *joins* para inserir duas tabelas em uma única dimensão.

Inicialmente levou-se em conta a conferência de dados nulos e com inconsistências no banco de dados, como cadastros de clientes incompletos e datas não registradas. No caso da dimensão cliente, a composição da mesma em primeira instancia deveria conter dados de endereço, porém como observou-se a falta de informação de alguns campos no banco de dados removeu-se esses atributos. Além disso, necessitou-se uma transformação complexa da data para a carga na dimensão tempo. A data contida no banco de dados da organização era uma única *string* e necessitava-se separá-la em dia, mês e ano, e após isso, transformando dias em semanas, para obtenção de informações semanais. A Figura 19 mostra a transformação

juntamente com a carga dos dados na dimensão cliente. E a Figura 20, mostra a mensagem de sucesso da carga.

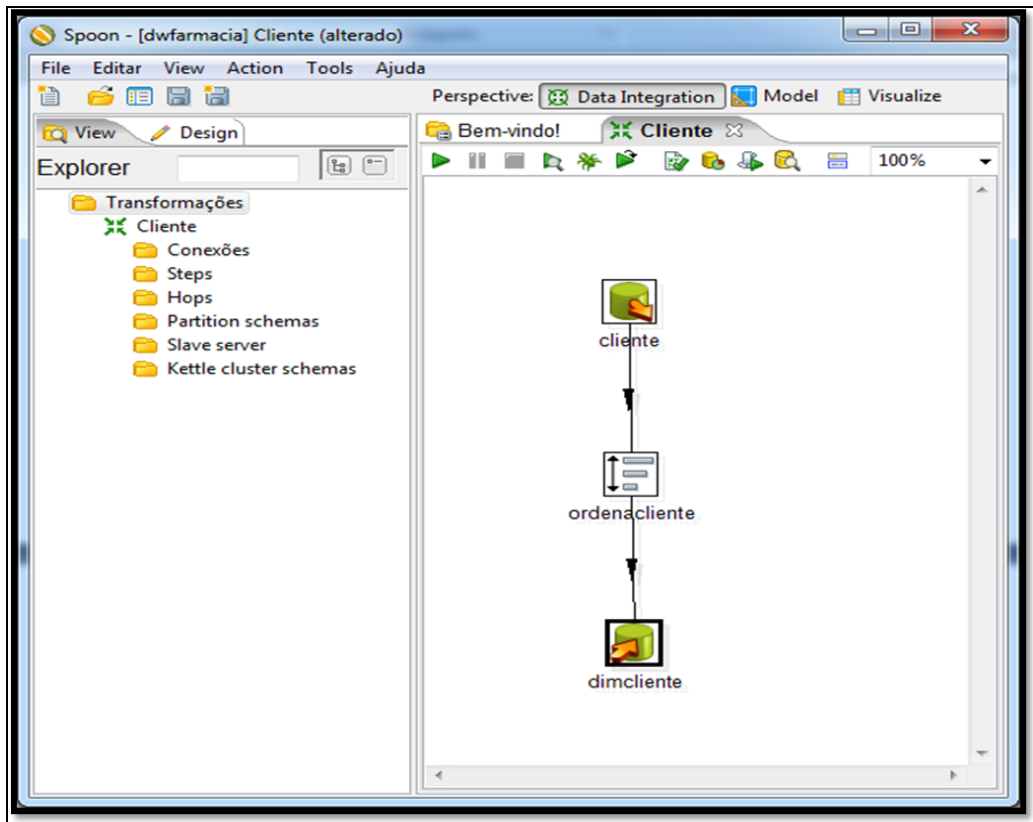


Figura 19 - Transformação cliente
Fonte: Adaptado da *Spoon*.

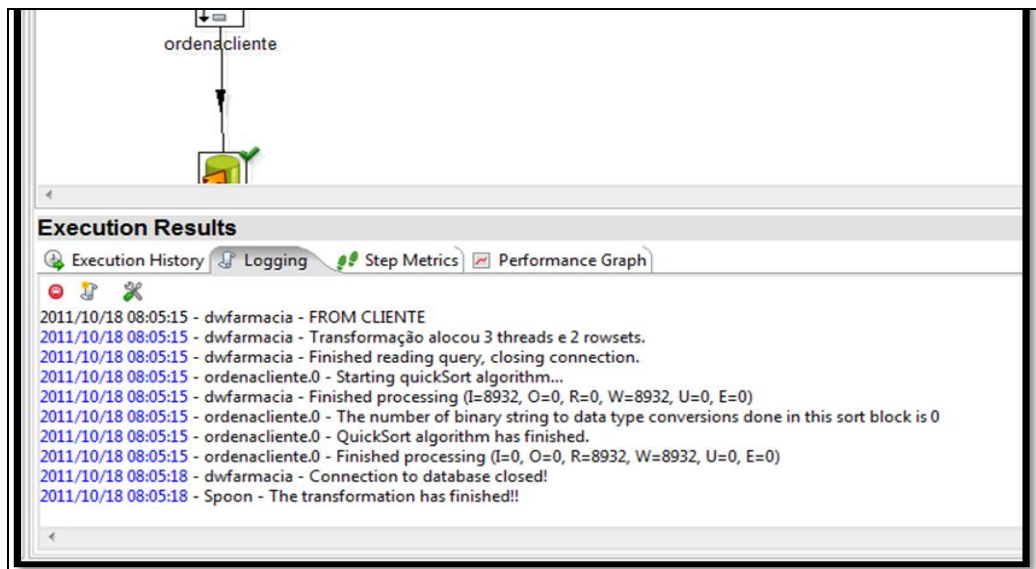


Figura 20 – Mensagem de sucesso da transformação cliente
Fonte: Adaptado da *Spoon*.

Pode-se traduzir as transformações e cargas feitas através da ferramenta *Spoon* em linguagem SQL. Como observa-se na Figura 21, a dimensão médico e a dimensão especialidade foram transformadas e carregadas em um mesmo processo. Observa-se a *query*

para a carga da dimensão médico. O primeiro passo é selecionar da tabela de origem aqueles campos que serão utilizados na tabela destino; no caso apresentado é o código do médico para ser utilizado como o campo identificador e auto-incremento, o nome do médico, seu número de registro e também seu código de especialidade. No Quadro 5 abaixo observa-se a explicação de cada cláusula SQL utilizada.

SQL	EXPLICAÇÃO
<pre>select m.cod_medico as id_medico, m.nome as nome_medico, m.nro_registro as num_registro, m.cod_especialidade as id_especialidade</pre>	<p>Seleciona-se os campos necessários para a dimensão medico.</p>
<pre>from medico m</pre>	<p>Refere-se à tabela de origem medico.</p>
<pre>order by m.cod_medico</pre>	<p>Ordena-se a consulta pelo código do medico.</p>

Quadro 5 – Explicação do comando SQL utilizado na transformação médico
Fonte: Autoria Própria.

O passo de ordenação observado na Figura 21 assume a sintaxe do *order by* no código da linguagem apresentada. E por fim, faz-se o *insert* na tabela de destino, como representado a seguir.

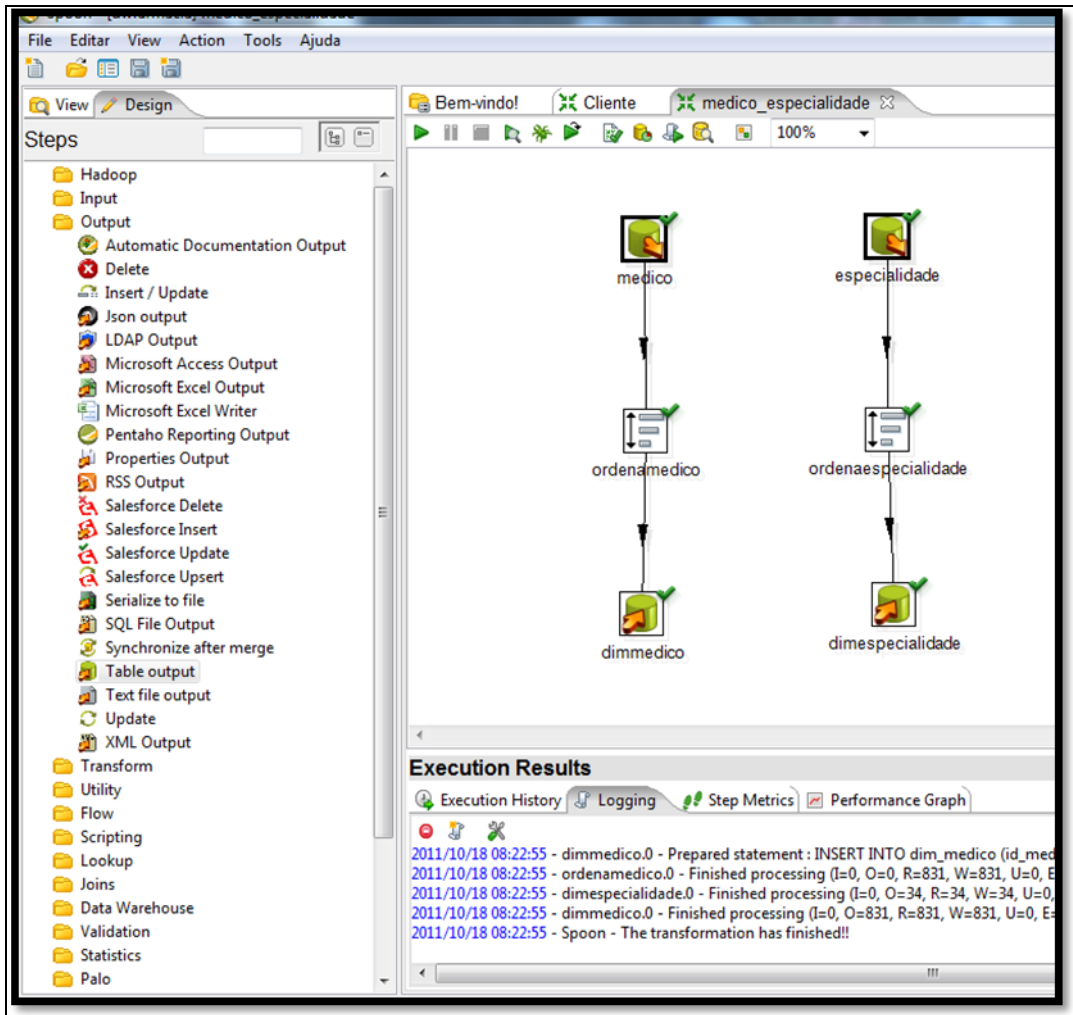


Figura 21 - Transformação médico e especialidade juntamente com a mensagem de sucesso
Fonte: Adaptado da Spoon.

Na dimensão venda foi necessário acrescentar o *mergejoin* para unir duas tabelas sendo elas: venda e itens_vendas. Nesse processo também foram ordenados os dados antes de realizar o processo de carga, podendo ser visualizado na Figura 22.

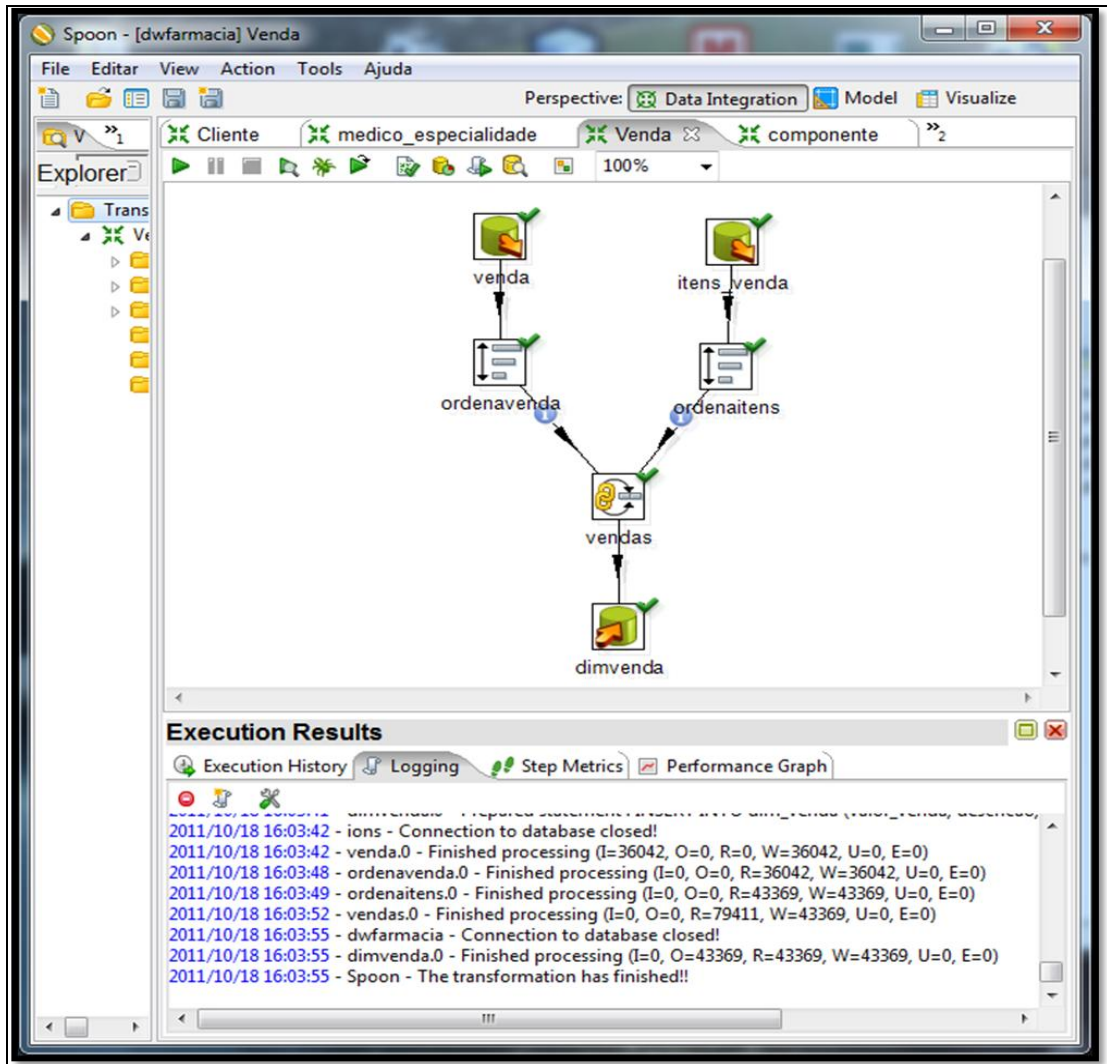


Figura 22 - Transformação Venda e mensagem de sucesso
Fonte: Adaptado da Spoon.

Além da dimensão venda tem-se uma simples carga e transformação da dimensão componente, visualizada na Figura 23. Na Figura 24, observa-se a transformação e carga da dimensão produto, que utiliza o *mergejoin* entre as tabelas Produto e Itens_Form_Prod.

Spoon - [dwfarmacia] componente

Perspective: Data Integration Model Visualize

Steps

- Hadoop
- Input
- Output
- Transform
- Utility
- Flow
- Scripting
- Lookup
- Joins
- Data Warehouse
- Validation
- Statistics
- Palo
- Job
- Mapping
- Bulk loading
- Inline
- Experimental
- Deprecated
- História

Execution Results

#	Nome do step	Copia nr	Lidos	escritos	Entrac
1	componente	0	0	968	96
2	ordenacomponentes	0	968	968	
3	dimcomponente	0	968	968	

Figura 23 - Transformação Componente e mensagem de sucesso
 Fonte: Adaptado do Spoon.

Spoon - [dwfarmacia] produto

Perspective: Data Integration Model Visualize

Steps

- Hadoop
- Input
- Output
- Automatic Doc
- Delete
- Insert / Update
- Json output
- LDAP Output
- Microsoft Acce
- Microsoft Excel
- Microsoft Excel
- Pentaho Report
- Properties Outp
- RSS Output
- Salesforce Dele
- Salesforce Inser
- Salesforce Upd
- Salesforce Upse
- Serialize to file
- SQL File Output
- Synchronize aft
- Table output
- Text file output
- Update
- XML Output

Diagram description: The diagram shows a data flow starting from 'produto' and 'itens_formula_prod'. 'produto' flows into 'ordenaproduto', and 'itens_formula_prod' flows into 'ordenaitens'. Both 'ordenaproduto' and 'ordenaitens' feed into a 'Merge Join' step, which then outputs to 'dimproduto'.

Figura 24 - Transformação Produto
 Fonte: Adaptado do Spoon.

A dimensão tempo passa por um processo diferente de transformação, pois é necessário dividir a *String* data em semana, mês e ano. Conforme ilustrado na Figura 25, é criado o *id* da dimensão tempo, após são selecionados os valores para então passar pelo passo de quebrar o tempo, isto é, dividir a data. Por fim, os campos são selecionados e inseridos na nova tabela.

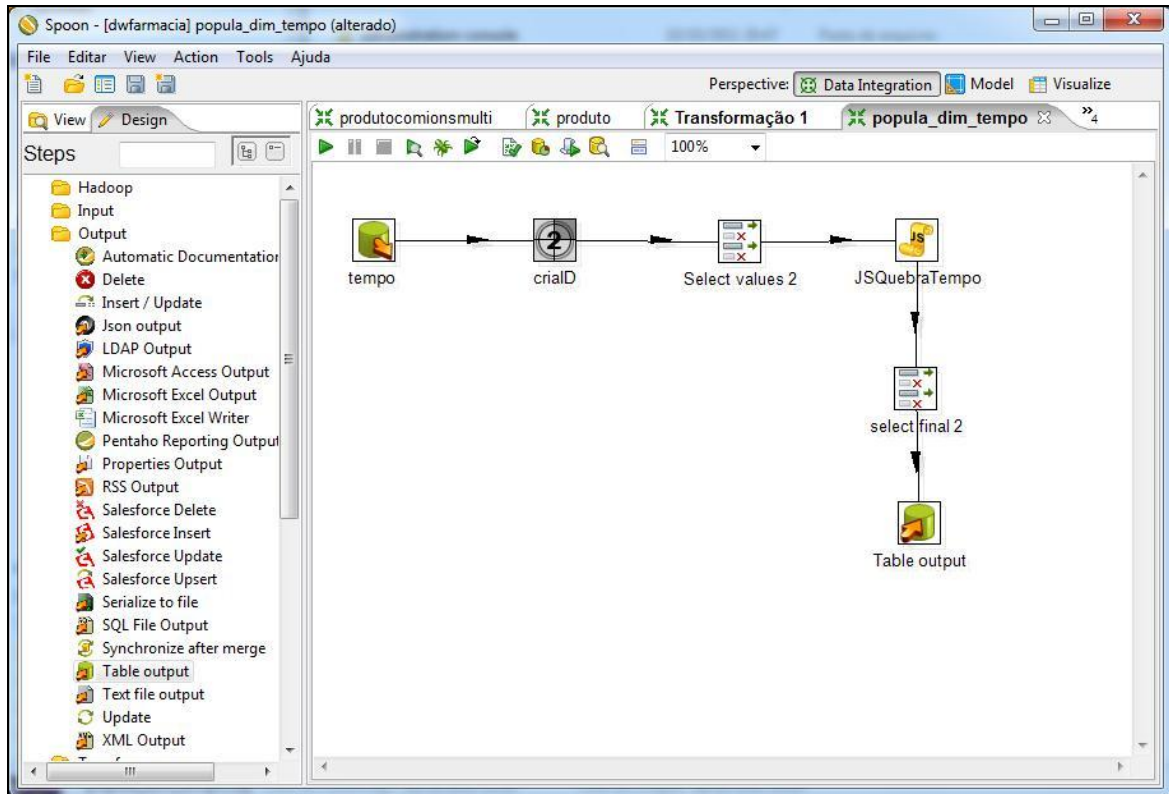


Figura 25 - Transformação Tempo

Fonte: Adaptado da *Spoon*.

correção da ferramenta aliando o próprio usuário que está dia a dia lidando com o *software* e sabe a real dificuldade encontrada.

Portanto a *Pentaho* juntamente com *iReport*, trazem consigo um novo conceito, representando a força da utilização de um *Software Livre*, deixando de lado um pouco crenças e rigidez e sim descobrindo a flexibilidade e poder de expansão e modificação desse novo modo de “pensar”.

5 IREPORT

Neste capítulo é descrito todo o processo da transformação da informação tácita em tangível, isto é, o desenvolvimento do relatório utilizando a ferramenta *open source iReport*. A Seção 5.1 debate sobre os principais motivos para a escolha da ferramenta *iReport* para utilização do estudo de caso. A Seção 5.2 apresenta conceitos importantes sobre as tecnologias que são utilizadas na ferramenta. Na Seção 5.3 apresentam-se alguns requisitos essenciais para o desenvolvimento de relatórios. Já na Seção 5.4 é prestado o ambiente da ferramenta. E na Seção 5.5 são apresentados os relatórios desenvolvidos para o estudo de caso

5.1 IREPORT E PENTAHO

Com o estudo aprofundado sobre a ferramenta Pentaho, percebeu-se a integração total da ferramenta escolhida para a realização do DW com a ferramenta *iReport*, escolhida para a construção dos relatórios. A opção pelo *iReport*, apesar de fortalecida pela integração entre as ferramentas, passou por um processo de comparação entre ela e a ferramenta nativa do Pentaho para geração de relatórios, mostrando-se superior em aspectos como: facilidade de aprendizagem e utilização, número de componentes inúmeras vezes maior, interface amigável com o desenvolvedor e trabalho adaptável de *query*.

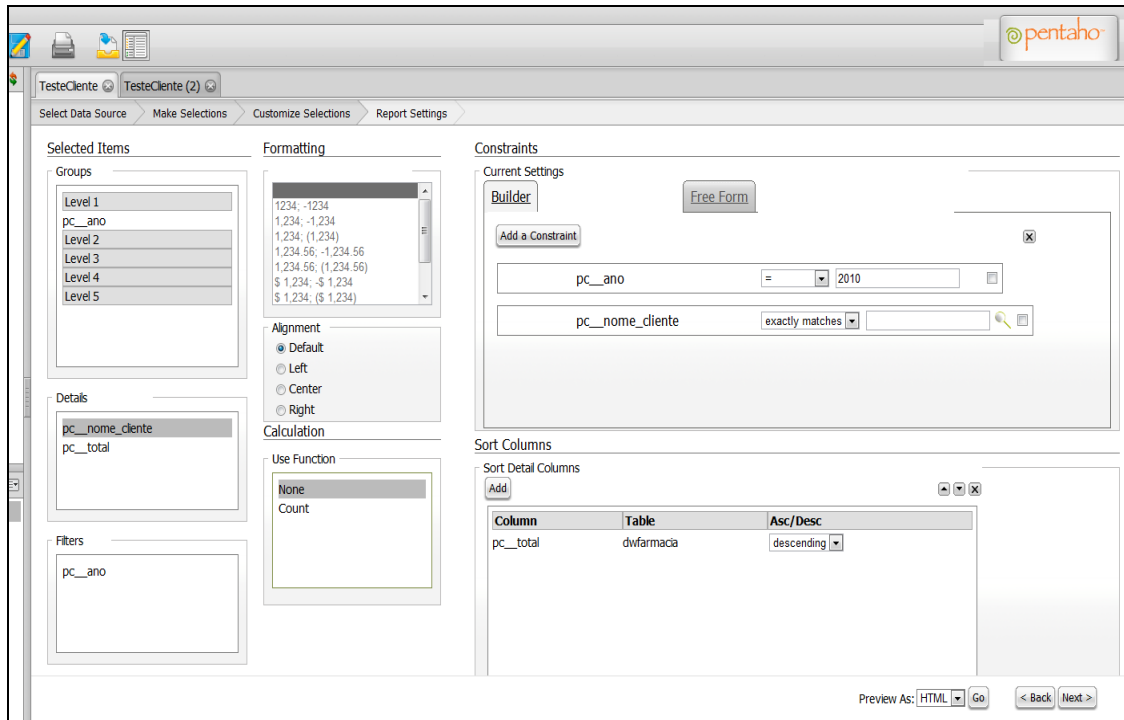


Figure 27 - Ambiente de desenvolvimento de relatórios da ferramenta Pentaho

Fonte: Adaptado da Pentaho

A ferramenta nativa, por focar o desenvolvimento rápido e de relatórios simples, deixa a desejar pela falta de componentes como *crostab*, *subrelatórios* e *tabelas*. Um mesmo relatório, desenvolvido com ambas as ferramentas, exigiu maior tempo de interpretação pelo usuário quando desenvolvido com a ferramenta nativa. No *iReport* foi desenvolvido utilizando-se o componente *crostab* – componente para referência cruzada – que facilita tanto o seu desenvolvimento, quanto sua interpretação. Na Figura 27 acima observa-se o ambiente de desenvolvimento da ferramenta nativa em que a configuração dos campos mostra-se confusa, dificultando o manuseio.

5.2 DEFINIÇÕES

Faz-se necessário entender algumas definições antes de partir para o desenvolvimento do relatório. Isto para uma melhor compreensão da ferramenta e para facilitar a sua utilização.

5.2.1 JasperReports

JasperReports é uma biblioteca *Java* que permite definir um relatório e depois executá-lo contra uma fonte de dados. A definição é feita em *XML* e pode ser editada manualmente, contudo utilizam-se geralmente ferramentas gráficas para sua edição. O

arquivo *XML* é compilado num arquivo com extensão *.jasper*, que é um formato binário, um meio mais rápido que o *XML* para a execução dos relatórios. O arquivo *.jasper* contém um arcabouço do relatório, porém destituído de dados. O arquivo código fonte em *XML* do *JasperReports* é o *.jrxml*.

A biblioteca permite a visualização dos relatórios em tela ou em arquivo *.PDF*, que junto com a facilidade de uso e sem qualquer custo das ferramentas, faz do *JasperReports* uma biblioteca de geração de relatórios completa.

5.2.2 JDK

Java SE Development Kit (JDK) +é um Kit de Desenvolvimento *Java*, ou seja, um conjunto de utilitários que permitem criar sistemas de *software* para plataforma *Java*. Contém todo o ambiente necessário para a criação e execução de aplicações *Java*, incluindo a máquina virtual *Java (JVM)*, o compilador *Java*, *APIs* do *Java* e outras ferramentas utilitárias.

5.2.3 *iReport*

O *iReport* é um programa *open source* – seus códigos fontes são distribuídos gratuitamente de acordo com a GNU (*General Public License*) – capaz de criar visualmente os mais complexos relatórios para aplicações *Java* no formato da biblioteca *JasperReports*. É desenvolvido na linguagem *Java*, o que significa que é multiplataforma.

Através de uma interface gráfica intuitiva, o desenvolvedor é capaz de criar qualquer tipo de relatório de forma simples e rápida. Mesmo sabendo que o *iReport* desenvolve um formato *XML* usado pelo *JasperReports* – o que não é difícil de manipular – para os desenvolvedores iniciantes o *iReport* evita a necessidade de modificações no código fonte. Para desenvolvedores experientes na manipulação desse formato, o tempo de desenvolvimento é reduzido consideravelmente com a ferramenta.

5.2.4 *Jar*

Java Archive (JAR) é um arquivo compactado usado para distribuir um conjunto de classes *Java*. É usado para armazenar classes compiladas e metadados associados que podem constituir um programa.

5.3 REQUISITOS PARA DESENVOLVIMENTO

A busca por *software* livre ou gratuito é um conceito de extrema importância no mundo da computação, que procura por ferramentas *freeware* (*software* gratuito), visando a economia por não necessitar da compra de licenças. Foi este um dos motivos pelo qual se fez a escolha destas ferramentas abaixo relacionadas. Das ferramentas citadas, faz-se a instalação apenas do *iReport*, disponível para consulta no Apêndice A, e todos os *downloads* das ferramentas podem ser feitos através do site dos mantenedores.

- **JDK** é o Kit de Desenvolvimento *Java* na versão 6.
- **jconn3** - *Driver* de conexão para o banco de dados.
- **iReport** - Ferramenta *iReport* na versão 3.7.3. Todo *.jasper* gerado nessa versão, é compatível à apenas essa versão. O *iReport* é a ferramenta de comunicação visual mais popular para o *JasperReports* (biblioteca de relatórios *Java*) e *JasperServer* (Reportagem de servidor).

5.4 O AMBIENTE IREPORT

5.4.1 Criando o *Template*

O primeiro passo é definir o *template* que será usado. Geralmente escolhe-se o padrão A4 que representa a folha de papel ofício. Para isso clica-se na opção do *menu* Arquivo e após em *New*, como mostra a Figura 28.

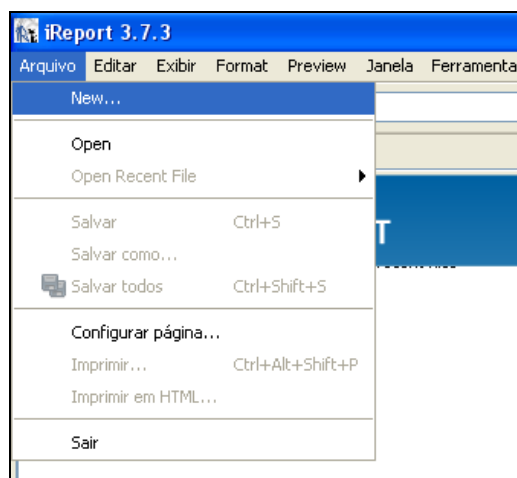


Figura 28 - Menu Arquivo
Fonte: iReport.

A tela de diálogo que aparece, seleciona a opção *Report*, escolhe-se *Blank A4*, como mostra a Figura 29, e clica no botão *Open this Template*.

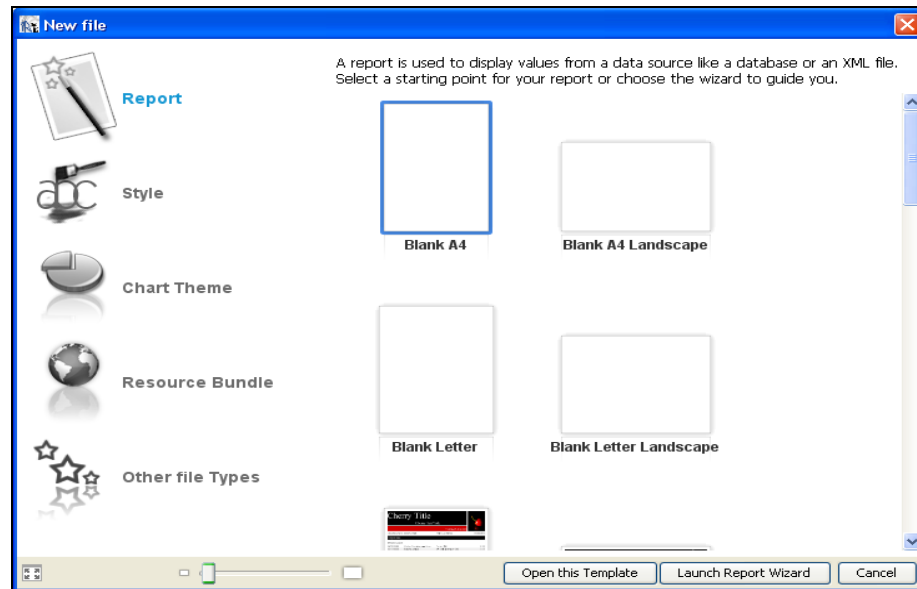


Figura 29 - New File
Fonte: *iReport*.

Escolhe-se o nome do relatório e o diretório onde deseja salvar e clica-se em próximo. Será mostrada uma tela de confirmação de criação do *template* e em seguida, a tela de abertura do *iReport*, como demonstrado na Figura 30 e Figura 31, respectivamente.

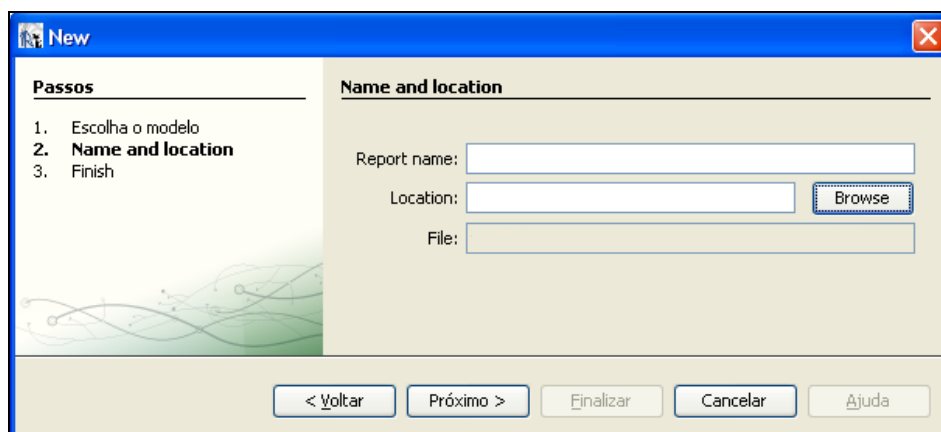


Figura 30 - Local para salvar o relatório
Fonte: *iReport*.

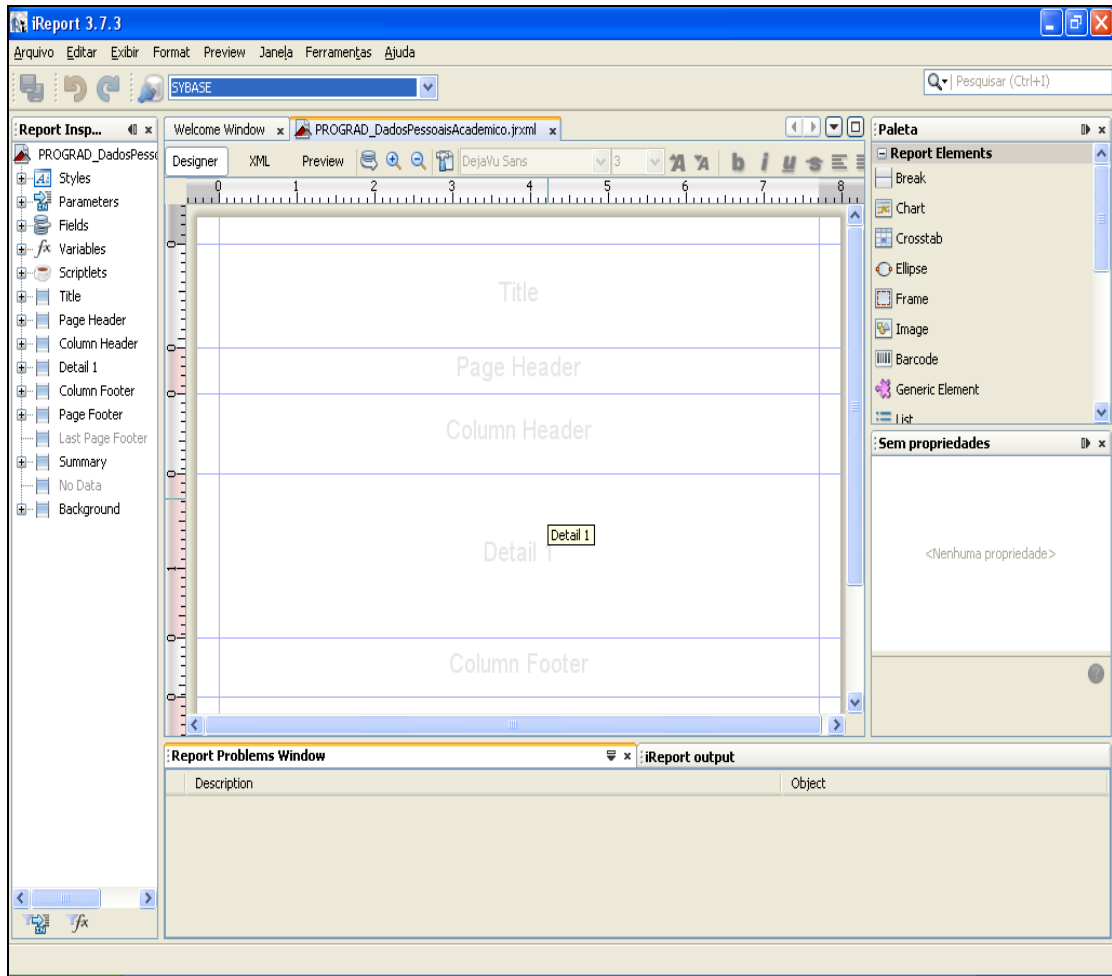


Figura 31 - Template Principal
Fonte: iReport.

Faz-se necessário excluir algumas linhas – *band* – restando apenas as linhas *PageHeader*, *Column Header*, *Detail* e *Page Footer*. Para isso faça como mostra a Figura 32, na paleta *Report Inspector*.

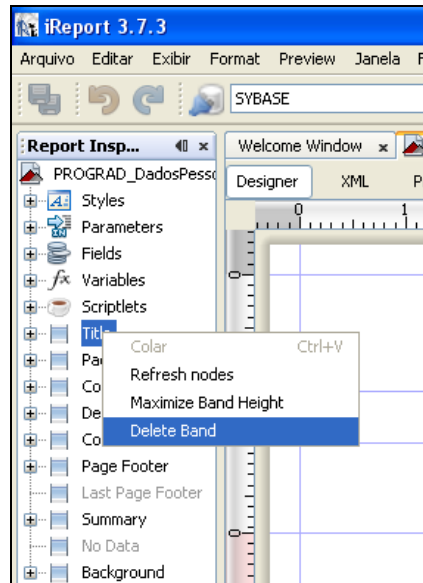


Figura 32 - Exclusão de título
 Fonte: *iReport*.

Após excluir as linhas que não serão utilizadas, a tela ficará como a Figura 33, notando que as linhas que foram excluídas ficam desabilitadas na paleta *Report Inspector*.

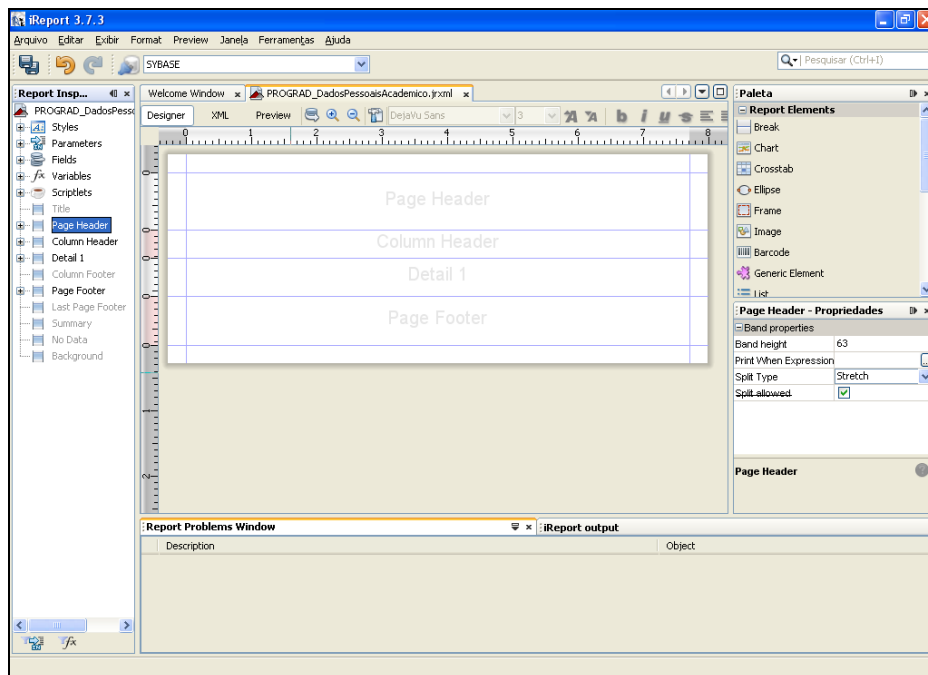


Figura 33 - Template Inicial mostrando linhas excluídas
 Fonte: *iReport*.

5.4.2 Conhecendo os componentes

Antes de partir para o desenvolvimento do relatório, faz-se necessário conhecer algumas das ferramentas disponíveis para o desenvolvedor, como a paleta *Reports Elements*, que possui muitos elementos que serão adicionados no relatório, como mostra a Figura 34:

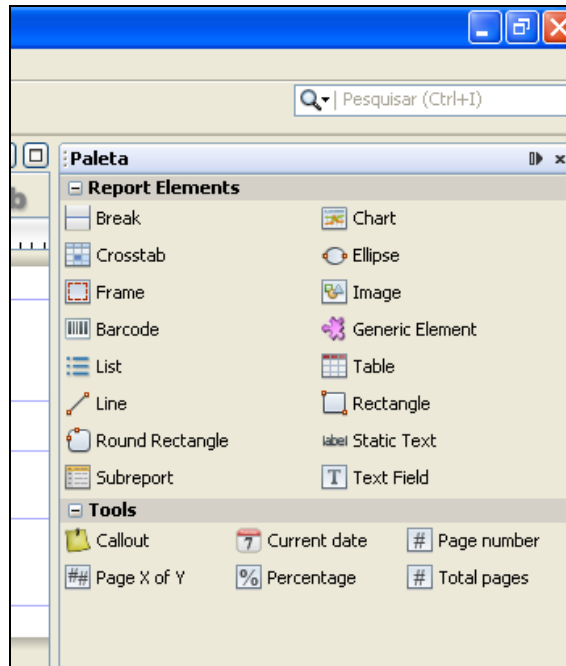


Figura 34 - Palheta
Fonte: *iReport*.

Para adicionar componentes ao relatório, basta clicar e arrastar os componentes para o local onde deseja. Como mostra a Figura 35, lembrando que as linhas (*bands*) que estão no relatório possuem funções específicas.

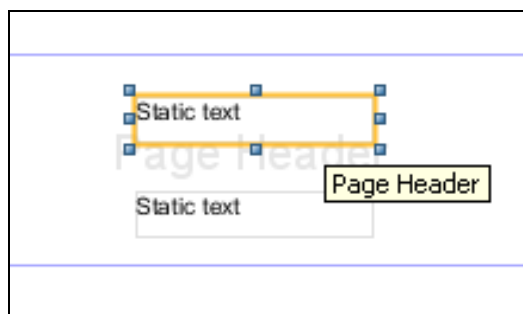


Figura 35 - Adicionar componente
Fonte: *iReport*.

PageHeader: Onde ficará geralmente dados estáticos, como o cabeçalho do relatório.

Column Header: Onde ficarão, geralmente, dados estáticos e que podem mudar de uma página para outra, como o título de uma tabela.

Detail: Onde trará os dados não estáticos da consulta SQL, são todas as linhas que aparecerão no relatório.

Page Footer: Onde pode haver geralmente dados estáticos, esta linha repetirá em todas as folhas do relatório, os dados exibidos geralmente são: da entidade que gerou o relatório, data e usuário.

5.4.3 A Aba *Tools*

As opções da aba *Tools* geralmente são utilizadas no *Page Footer*, como por exemplo, o *Current date* e o *Page X of Y*, dependendo da necessidade do relatório.

Cada componente tem suas propriedades, como é o caso da data (*Current date*), que ao adicionar o campo escolhe-se o formato.

5.4.4 Palheta de Componentes

- *Chart* – Componente para adicionar gráficos nos relatórios, existem diversos formatos, escolhendo o modelo que irá melhor atendê-lo, lembrando que cada gráfico tem um jeito de utilizar.

- *Crosstab* – Componente que permite criar uma tabela com referência cruzada.

- *Ellipse* – Componente visual, geralmente utilizado para melhorar o *layout*.

- *Frame* – Componente utilizado para agrupar outros componentes.

- *Image* – Componente visual para inserir uma imagem, esta imagem precisa estar onde o relatório será gerado.

- *List* – Permite criar uma nova fonte de dados e apresentá-la como se fosse um *Band Detail*.

- *Table* – Cria uma tabela simples.

- *Line* – Componente visual, geralmente utilizado para melhorar o *layout*.

- *Rectangle* – Componente visual, geralmente utilizado para melhorar o *layout*.

- *Round Rectangle* – Componente visual, geralmente utilizado para melhorar o *layout*.

- *Static Text* – Componente visual, geralmente utilizado para escrever algo estático no relatório, como títulos.

- *Subreport* – Componente muito utilizado para agrupar relatórios, agrupar relatórios diferentes (como se fosse *master* e *slave*), conseguir uma maior flexibilidade em um determinado relatório.

- *Text Field* – Componente utilizado como um campo de retorno de uma consulta *SQL*.

5.4.5 Modelando o Relatório

A opção *Static text* é utilizada para textos fixos, como títulos, cabeçalhos e rodapés. Depois de adicionados os campos estáticos, criam-se os parâmetros. São eles os utilizados para filtrar os dados do relatório e são aqueles na cláusula *where* da *SQL*. Adiciona-se parâmetros na paleta *Report Inspector*, na opção *Parameters*, como mostra a Figura 36:

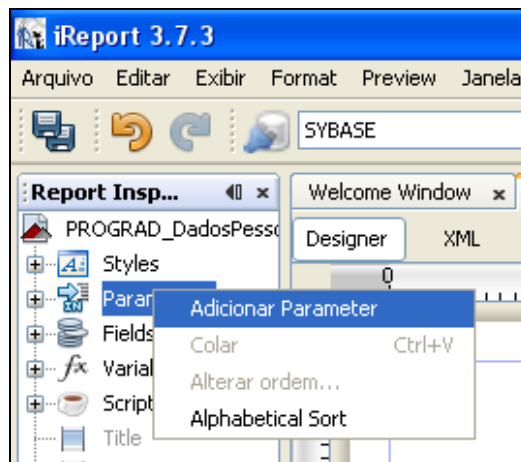


Figura 36 - Adicionar parâmetros
Fonte: iReport.

Para configurá-lo, utilizam-se as opções disponíveis na paleta Propriedades. Primeiramente define-se um valor *default* e o tipo de dado.

Depois de criados todos os parâmetros adicionam-se a *SQL* (*Structure Query Language*). Para adicionar uma *SQL*, seleciona a paleta *Report Inspector* (canto superior esquerdo), clica-se com o botão direito do *report name*, abrindo assim um *menu*, onde clica-se na opção *Edit Query*, abrirá uma tela como a Figura 37.

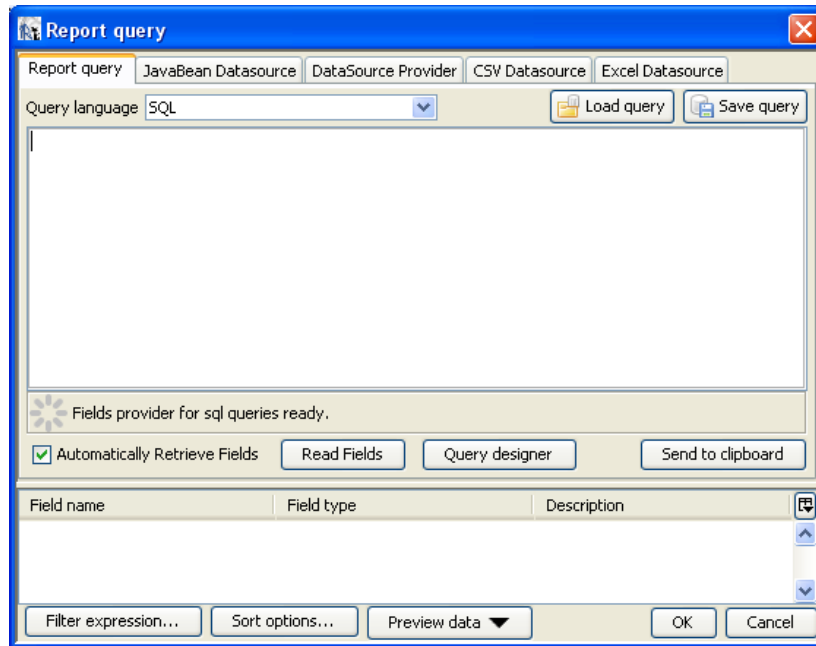


Figura 37 - Editar Query
Fonte: iReport.

Digita-se a *query* no primeiro bloco em branco que aparece na tela, semelhantemente a Figura 38:

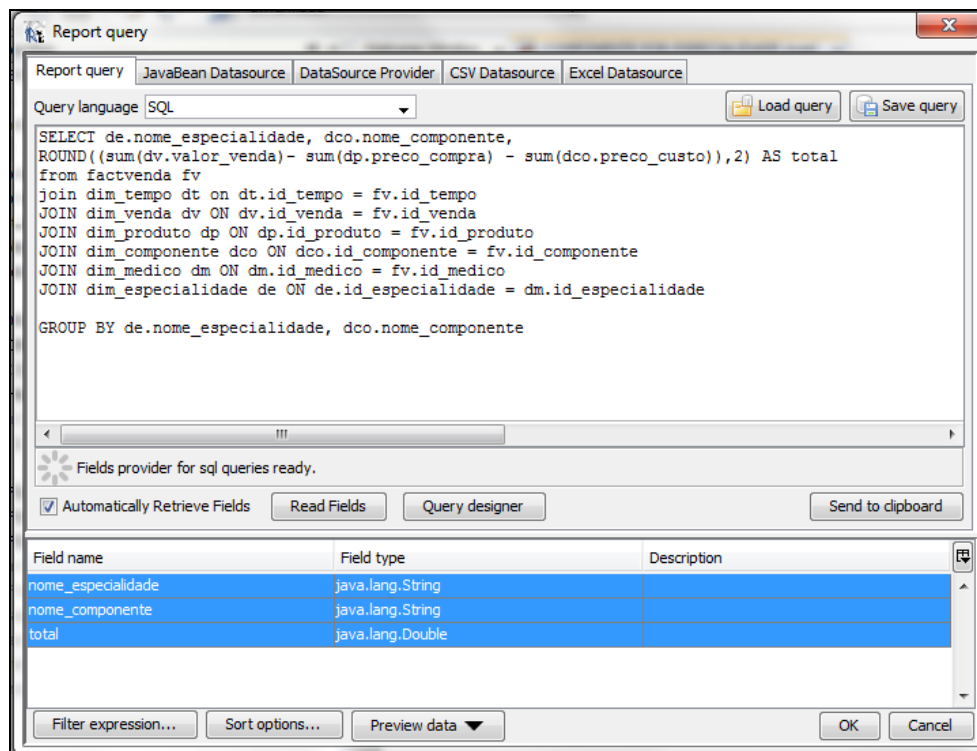


Figura 38 - Editar Query
Fonte: Adaptado do iReport.

Quando se digita a SQL, existe um verificador que testa se a *query* está correta. Em caso afirmativo, serão listados os *fields*, como mostrado na Figura 38. Caso ocorra algum

erro, o verificador avisará, mostrando mensagens no segundo bloco com o máximo de informações sobre o erro.

A SQL possui três pontos importantes: a cláusula *SELECT*, seguida dos campos onde terão os dados da consulta, o *FROM*, que são as tabelas usadas e o *WHERE*, é a parte mais importante nesse momento, pois é nessa cláusula que são inseridos os parâmetros para a execução da SQL.

Os parâmetros são passados da seguinte maneira ($\$P!\{\text{nome_do_parametro}\}$), devendo ser o mesmo nome que foi criado anteriormente. Por isso é importante dar nomes intuitivos e que tenham significado único.

Após a SQL terminada e sua validação concluída, os seus campos (os que serão buscados) aparecem na paleta *Object Inspector*, no submenu *Fields*, como mostra a Figura 39:

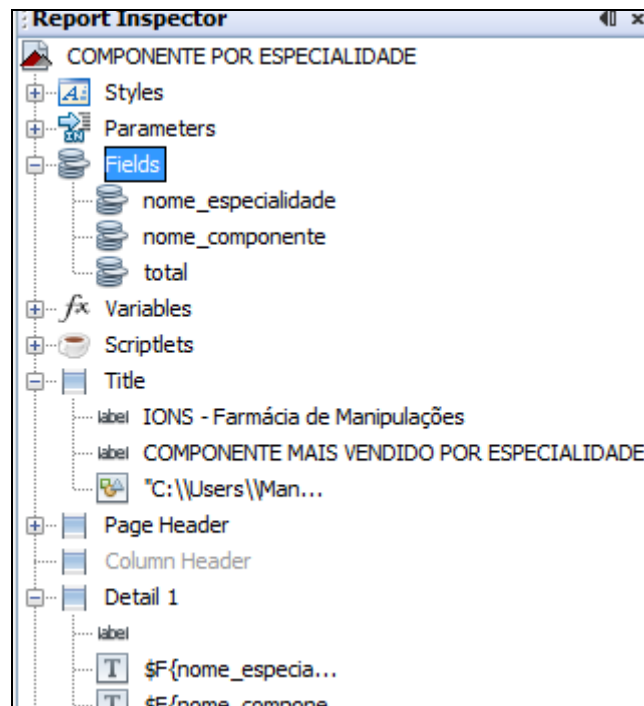


Figura 39 - Submenu *Fields*
Fonte: *iReport*.

Nesse instante, é iniciada a construção do relatório em si. Coloca-se os textos estáticos na linha do *Page Header* – cabeçalho – e coloca-se os *Fields* na linha do *Detail*(que são os dados retornados pela consulta SQL), como mostra a Figura 40:

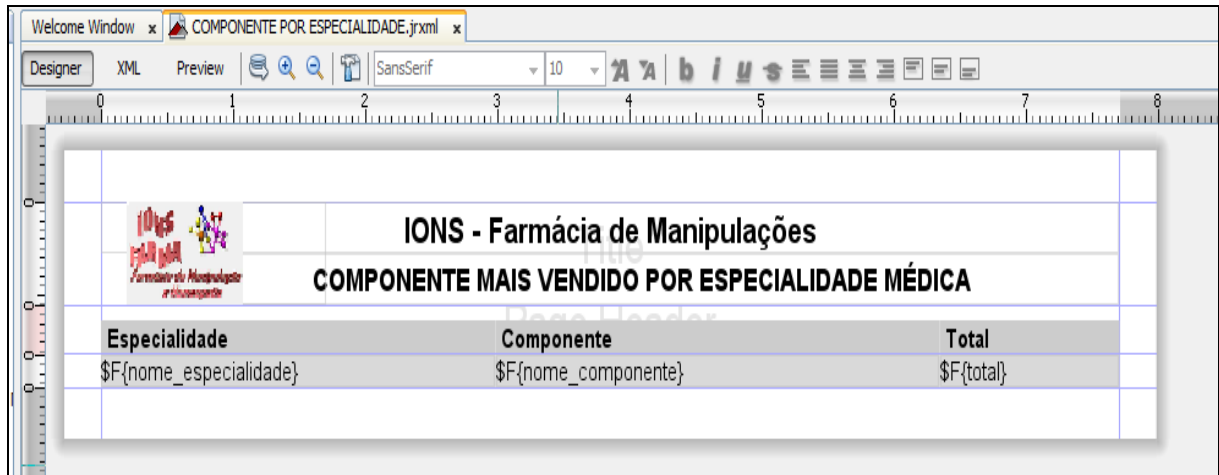


Figura 40 - Header
Fonte: Adaptado do *iReport*.

Ajusta-se o *layout* do relatório, podendo ser configuradas diversas propriedades como tamanho da fonte, altura do texto estático e dos campos, tamanho das linhas (*band*), comprimento, espaçamentos e bordas. Após todos os ajustes, testa-se o relatório através da opção *Preview*, como mostra a Figura 41:

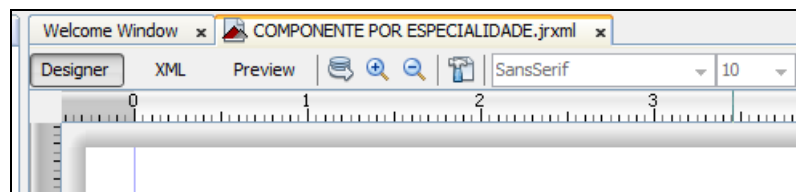


Figura 41 - Preview
Fonte: Adaptado do *iReport*.

Os parâmetros utilizados serão solicitados, podendo ser possível desabilitá-los através do *menu* de propriedades dos parâmetros, desmarcando a opção “*use as a prompt*”. Ao final da solicitação de parâmetros, aparecerá o relatório, como mostra a Figura 42:

IONS - Farmácia de Manipulações COMPONENTE MAIS VENDIDO POR ESPECIALIDADE MÉDICA		
Especialidade	Componente	Total
ALERGOLOGISTA/IMUNOLOGISTA	CREME LANETTE	16.44
ALERGOLOGISTA/IMUNOLOGISTA	DESONIDA	-2.77
CARDIOLOGISTA	A.A.S.	821.95
CARDIOLOGISTA	AGUA DESTILADA	298.09
CARDIOLOGISTA	ALOPURINOL	3191.42
CARDIOLOGISTA	AMILORIDA CLORIDRATO	92.79
CARDIOLOGISTA	AMIODARONA HCL	147.52
CARDIOLOGISTA	ANLODIPINA BESILATO	161.79
CARDIOLOGISTA	BIOFLAVONOIDES	128.26
CARDIOLOGISTA	CAFEINA	235.35
CARDIOLOGISTA	CALCIO DE OSTRAS	205.29
CARDIOLOGISTA	CARBONATO DE CALCIO	32.41

Figura 42 - Relatório
Fonte: Adaptado do *iReport*.

Quando visualiza-se o relatório, é gerado o arquivo *.jasper* (no mesmo diretório onde foi criado o relatório).

5.5 IMPLEMENTAÇÃO DO ESTUDO DE CASO

Todos os dados apresentados nos relatórios foram simulados.


 IONS - Farmácia de Manipulações FATURAMENTO POR MÉDICO			
	2009	2010	2011
ALCIONE FRANCISCO BARAUSSE	135.76	1085.53	422.76
ALEXANDRE BARAO ACUNA	203.28		
ALFREDO A. JUNIOR	0.0		
AMAURY R. NOGUEIRA JR	251.52		
AMILCAR RUANI	39.04	221.24	
ANA LORETE FRANK	44.18		
ANA PAULA CARVALHO		255.27	
ANITCIA ADENIA BUENO		11.6	
ANTONIO ALCIDES KLUG JR	0.0	15.81	1032.66
ARIEL GERALDO DE ALMEIDA	71.27		
AUREA LUCIA G. DE S. CAMPOREZ	2117.59		
BETINA FAGGION	154.66	464.55	21.97
CARLA KOMNITSKI		118.93	308.85
CARLOS ALBERTO BATISTA SILVA	101.45		
CARLOS ALBERTO STRACHEWSKI	572.86	2809.34	898.87
CAROLYN MARIA DE G WENCESLAU		380.54	1480.05
CASEMIRO PRZEPIURA	52192.71	272170.01	285747.6
CELSO HIPOLITO PILGER			0.0
CLAUDIA G. L. STURZENEKER	244.05		
CLAUDIA NEKATSCHALOW	77.92		
DANIELA MATTAR		10.97	
DAVID MAURICIO COUTINHO	115349.62	123554.58	96692.04
DOMINGOS ALUISIO CAMPOREZ	0.0	191.2	
EDAR GERTNER			88.92
EDISON MATOS DE OLIVEIRA		10.1	
EDSON MAFFEZZOLLI			1512.7
ELAINE PRANDEL		241.51	
ELIANE CRISTINE GALVAO			1038.01
ELOINA DO ROCIO VALENGA BARONI	28.22	423.34	

Figura 43 - Relatório Médico Faturamento
 Fonte: Adaptado do *iReport*.

O primeiro relatório gerado, Figura 43, responde a primeira pergunta sobre qual é o faturamento por médico nos anos de 2009, 2010, 2011.

O segundo relatório, representado pela Figura 44, demonstra o faturamento do cliente nos anos de 2009, 2010, 2011.


 IONS - Farmácia de Manipulações FATURAMENTO POR CLIENTE			
	2009	2010	2011
Adalgisa Maria Ribeiro Bachosk		2928.91	
Adelis Barauce			2058.37
Adriana Bellanda		19.97	
Adriana Hilbert		22.06	
Adriana Terezinha Sukoski		204.18	
Adryelli C. Capoia	19.2		
Agenor Gambassi	30.66	15.81	
Aida dos Santos Lima			27.96
Aladia Mottin		551.16	
Alberto Strachewski	2694.42	2809.34	1379.34
Alcione Maria Pedroso			3135.2
Alda Aparecida de Moraes	2083.21	26.86	
Alda da Silva Oliveira	117.55		
Aleisa Schmid		20.63	1344.35
Alencar Ulisses de C. Rizental		19.86	
Alessandra Sovinski			3766.11
Alex Lourenço		2.02	
Alex Monteiro Vedan	182.74		
Alexandre C. Gmyterco		19.97	
Alexandre Lunardon	19.86		
Alfredo Madureira	-23.32		
Alice Barth da Silva	100.12	20.12	20.12
Aline Bochnia			43.33
Aline C. Oliveira			748.55
Allyson Alberto Wsselevicz	19.2		
Almir Messias			614.67
Aloisio Sukoski		953.48	
Altivir Martins			1189.85
Amanda Baraus Bacila			3492.98

Figura 44- Relatório Faturamento Cliente
Fonte: Adaptado do iReport.

O terceiro relatório demonstra o faturamento por componentes nos anos de 2009, 2010, 2011, como pode ser observado na Figura 45.


 IONS - Farmácia de Manipulações FATURAMENTO POR COMPONENTE			
	2009	2010	2011
5 OH TRIPTOFANO	323.4	263.76	50.89
A.A.S.	10.19	228.17	91.43
ABACATEIRO PO		1233.68	
AC. NITRICUM CH2	19.86		
ACEROLA EXTRATO		20.49	
ACIDO ALFA LIPOICO	1375.28	1848.42	
ACIDO CITRICO		404.5	
ACIDO FOLICO	100.96	105.78	14.64
ACIDO GLICOLICO		206.03	
ACIDO GLUTAMICO	668.81	198.04	803.11
ACIDO KOGICO		89.54	
ACIDO MANDELICO		42.81	
ACIDO SALICILICO	37.98	56.63	38.88
ACONITUM CH2	102.19	20.63	
ADVANTRA	159.6		
AGAR AGAR EXT.	687.99	1288.13	334.18
AGARICUS	57.24	114.71	180.56
AGRIAO TINTURA			191.25
AGRIMONY			0.0
AGUA DESTILADA	0.0	49.12	19.26
ALANTOINA	271.94	3753.66	1078.43
ALCACHOFRA EXTRATO SECO	150.55	26.04	1526.91
ALCOOL DE CEREAIS		144.65	
ALENDRONATO DE SODIO	34.11	366.54	257.23
ALFA BISABOOL		27.71	32.31
ALHO 250MG - CAPS.OLEOSA	129.07	32.4	
ALLIUM CEPA SD1		0.26	68.11
ALLIUM CEPA CH2	0.99		
ALOINA		3999.28	2909.15

Figura 45 - Relatório Faturamento Componente
 Fonte: Adaptado do *iReport*.

O quarto relatório, apresenta o faturamento por especialidades médicas, representado pela Figura 46.


		IONS - Farmácia de Manipulações		
		FATURAMENTO POR ESPECIALIDADE MÉDICA		
	2009	2010	2011	
ALERGOLOGISTA/IMUNOLOGISTA	13.67			
CARDIOLOGISTA	5656.84	8799.97	14214.99	
CIRURGIAO PLASTICO	1.06			
CIRURGIAO VASCULAR		135.36		
CLINICO GERAL	36695.18	37319.17	37942.94	
DENTISTA	111.74			
DERMATOLOGISTA	6743.29	3054.8	442.63	
ENDOCRINOLOGISTA	224.26	-9.35		
GASTROENTEROLOGISTA	502.99	9.52	3878.07	
GERIATRA	159.6	289.88	236.85	
GINECOLOGISTA	771.4	1178.73	211.12	
HOMEOPATA	0.19	243.21		
NEUROLOGISTA	8.48	69.54		
ORTOMOLECULAR	115349.62	123554.58	100458.15	
ORTOPEDISTA	145.64	1283.5	455.67	
PEDIATRA		10.1		
PSIQUIATRA	149.62	107.06		
REUMATOLOGISTA		474.77	1480.05	
TERAPEUTA	52192.71	272170.01	285747.6	
UROLOGISTA	39.04	244.88		
VETERINARIO	554.45	93.6	161.47	

Figura 46 - Relatório médico por especialidade
Fonte: Adaptado do *iReport*.

O quinto relatório indica os meses de maior e menor faturamento respectivo da organização no corrente ano, indicado na Figura 47.


 IONS - Farmácia de Manipulações MÊS DE MAIOR FATURAMENTO E MÊS DE MENOR FATURAMENTO	
Mês	Total
Junho	81416.28
Setembro	15712.31

Figura 47 - Relatório maior e menor faturamento do ano 2011

Fonte: Adaptado do *iReport*.

O sexto relatório apresenta o total de faturamento nos anos de 2007, 2008, 2009, 2010 e 2011, indicado na Figura 48.


 IONS - Farmácia de Manipulações TOTAL DO FATURAMENTO POR ANO	
ANO	TOTAL
2007	68802.74
2008	115077.36
2009	219319.76
2010	449009.33
2011	445229.55

Figura 48- Relatório total de faturamento para os anos 2007 a 2011

Fonte: Adaptado do *iReport*.

6 CONCLUSÃO

Após a fase de estudo da tecnologia, é possível concluir que a melhor implementação do *Data Warehouse* está diretamente relacionada com o entendimento e compreensão de cada fase do processo, bem como associá-las com as ferramentas. Caso o analista não esteja familiarizado com o seu processo na teoria, dificilmente conseguirá colocá-lo em prática e assim, atingir suas metas, sendo fadado a refazê-lo.

Na implementação do *Data Warehouse* e do estudo de caso fez-se necessário a integração total do analista com o processo empresarial, uma vez que cada organização tem suas particularidades. No caso apresentado, esse estudo do ambiente corporativo foi possível graças ao Modelo de Entidade e Relacionamento disponibilizado pela empresa, que teve de ser minuciosamente compreendido além da realização do *brainstorming* para obter questões a serem resolvidas com o estudo.

A ferramenta utilizada no desenvolvimento, que por ser acessível e permitir se adaptar a casos específicos é pouco acadêmica, isto é, de difícil aprendizagem. Segue rotinas ortodoxas, logo, quando configurada de maneira ambígua, não permite prosseguir o processo. Isso faz com que o desenvolvimento tenha de ser reiniciado diversas vezes, pois se exige pré-requisitos até então desconhecidos, diminuindo a produtividade e prolongando o tempo em cada fase.

Vale ressaltar que a *Pentaho* engloba todas as tarefas do processo de *Business Intelligence* e ainda com a possibilidade de utilizar qualquer banco de dados. Com esta vasta gama de recursos, para se obter um bom resultado, depende de muito estudo de suas inúmeras funcionalidades.

Ao trabalhar com a ferramenta *open source iReport*, percebeu-se certa facilidade ao compor os relatórios. A ferramenta exige um bom conhecimento da linguagem SQL para retornar resultados eficientes. As configurações relativas ao *design* dependem do gosto do desenvolvedor, uma vez que possui muitos recursos. Nos casos apresentados o componente mais utilizado foi o *Crosstab*.

As respostas das questões levantadas na etapa de *brainstorming* foram respondidas com sucesso e com certa facilidade, podendo ser mais explorado em novos relatórios, dependendo das necessidades dos usuários.

Em geral, as expectativas com o desenvolvimento do DW foram alcançadas. O estudo da teoria foi essencial para realização dos processos e a busca pela ferramenta ideal fez

a equipe conhecer possibilidades distintas auxiliando para a realização do estudo de caso. A *Pentaho* supriu as necessidades para realizar os processos, apesar de toda a dificuldade para trabalhar com esta vasta gama de recursos. Os relatórios publicados no formato *.jasper* apresentaram integração total com a ferramenta de desenvolvimento, podendo ser considerada uma opção à ferramenta da suíte *Pentaho* utilizada para o mesmo fim.

Assim, considera-se que o objetivo de utilizar um caso prático para experimentar a utilização da tecnologia de *Data Warehouse* e *iReport* foi cumprido, apresentando a transformação do conhecimento tácito em tangível, e abrindo portas para trabalhos futuros.

7 TRABALHOS FUTUROS

Para trabalhos futuros, pode ser mencionada a aplicação de *Data Warehouse* para outros processos de negócio, já que o estudo desenvolvido neste trabalho foi relacionado somente à venda. Além disso, pode-se utilizar outra ferramenta para desenvolvê-lo, o que possibilita um trabalho de análise comparativa dos resultados obtidos entre as duas ferramentas.

Sugere-se também realizar processo de mineração de dados, utilizando-se do recurso de *Data Mining*, que procura buscar padrões em grandes quantidades de base de dados, como é o caso da base histórica do *Data Warehouse*.

REFERÊNCIAS

ANVISA - Agência Nacional de Vigilância Sanitária. Resolução RDC N° 214, de 9 de fevereiro de 2007. Brasília (DF): Diário Oficial [da] República Federativa do Brasil; 9 fev. 2007.

ARAÚJO, Erika M. T.; BATISTA, Mônica de L. S. **Qualidade na Modelagem dos Dados de um Data Warehouse**. Disponível em < <http://www.devmedia.com.br/post-6978-Qualidade-na-Modelagem-dos-Dados-de-um-Data-Warehouse.html>> Acesso em: 15 out 2011.

BATISTA, Marcelo S. **Data warehouse**. Faculdade Latino Americana, Anápolis, Goiás. Disponível em < https://docs.google.com/viewer?a=v&q=cache:njq6UnYXfmIJ:drc.objectis.net/faculdade-anhanguera-de-apolis/pra/banco-de-dados-ii/Artigo%2520descrevendo%2520o%2520que%2520201a%2520um%2520Data%2520warehouse.doc/at_download/file+artigos+sobre+data+warehouse&hl=pt-BR&gl=br&pid=bl&srcid=ADGEESgwRqyAXXR0dXwNmkIDwhz-ZwuvCDtGb5TdCs5DpoqZ7azXFPRFa8YPPdDAe2j3iDA0PyJEC5bLYCeKKunNoF7JsB0Z2ZflTodpeRuZgMazSSxIj3PKb8LXz6OdGzQDYe2_tmUz&sig=AHIEtbThhLLhSZcx22BiYPO2FoOtLfUohA> Acesso em : 14 out 2011.

BARQUIN, R.; EDELSTEIN, S. **Planning and Designing the Data Warehouse**. Prentice Hall, 1997.

CASSARO, A. C. **Sistemas de Informação para a Tomada de Decisão**. São Paulo: Pioneira Thomson Learning, 2003. 3. Reimpr da 3ed. rev. e ampl. De 1999.

COUGO, P. **Modelagem conceitual e projeto de bancos de dados**. Rio de Janeiro: Campus, 1997.

DAVENPORT, T. H.; PRUSAK, L. **Conhecimento Empresarial: como as organizações gerenciam seu capital intelectual**. Tradução de Lenke Peres – Rio de Janeiro: Elsevier, 2003.

DE SORDI, J. O. **Administração da informação: fundamentos e práticas para uma nova gestão do conhecimento**. São Paulo, Editora: Saraiva, 2008.

FAVARETTO, Fabio; RHODEN, Carlos A. **Considerações finais sobre atividades de identificação, localização e tratamento de dados na construção de um Data Warehouse.**In: Programa de Pós Graduação em Engenharia de Produção e Sistemas.Pontifícia Universidade Católica do Paraná,Curitiba. Disponível em <http://conged.deinfo.uepg.br/~iconged/Artigos/Artigo_05.pdf > Acesso em: 1ª out 2011.

FUJIWARA, D. K. **Data Warehouse Como Instrumento de Suporte à Avaliação Acadêmica.** Brasília: UnB, 2006.

GRAY, P; WATSON, H. J. **Decision Support in the Data Warehouse.** New Jersey: Prentice Hall, 1998.

GOLDSCHMIDT, R.; PASSOS, E. L. **Data Mining: um guia prático.** Rio de Janeiro: Elsevier, 2005.

HOKAMA, Daniele D,B; et al. **A Modelagem de Dados no ambiente Data Warehouse.** 2004. 110f. Monografia, Universidade Presbiteriana Mackenzie, 2004. Disponível em: <<http://meusite.mackenzie.com.br/rogerio/tgi/2004ModelagemDW.pdf> > Acesso em: 03 jun 2011.

INMON, W. H.; WELCH, J. D.; GLASSEY, K. L. **Gerenciando data warehouse.** São Paulo: Makron, 1999.

IREPORT, Disponível em <<http://jasperforge.org/project/ireport>>. Acesso em: 22 jul 2011.

KIMBALL, R. **Data Warehouse Toolkit.** Makron Books, Rio de Janeiro: Campus, 1998.

MACHADO, F. N R. **Tecnologia e Projeto de Data Warehouse Uma Visão Multidimensional.**São Paulo: Editora Érica, 2004.

NAVARRO, Cristina de A. **Data Warehouse.** Brasília, 1996. Disponível em: <http://www.serpro.gov.br/imprensa/publicacoes/tema-1/tematec/1996/ttec27>>. Acesso em: 22 jul 2011.

OLIVEIRA, D. P. R. **Sistemas de informações gerenciais: estratégicas, táticas, operacionais.** São Paulo: Atlas, 2005. 10 ed.

OLIVEIRA, W. J. **Data warehouse**. Florianópolis: Visual Books, 2002.

PENTAHO, Disponível em <<http://www.pentaho.com/>>. Acesso em: 14 set 2011.

RODRIGUES, Sander. C. **Análise Comparativa das Ferramentas OLAP e Pentaho, Analysis Services**. Centro Universitário de Belo Horizonte UNI-BH, 2010. Disponível em <https://docs.google.com/viewer?a=v&q=cache:yILNdrFEZy8J:https://www.jhcruvinel.com/index.php%3Foption%3Dcom_docman%26task%3Ddoc_view%26gid%3D70%26Itemid%3D2+Analise+Comparativa+das+Ferramentas+OLAP+e+Pentaho,+Analysis+Services.&hl=pt-BR&gl=br&pid=bl&srcid=ADGEESjRO9cTjOXxfuXxNfUJ7Oj57nA3Ln_p102T8G2VTSC5k7JgvFVtz3exW1p0X8ywdNUIBexe9C3rfZhSDE3CfB1eceyt4X0-Tl8pXuBu-31ZI8egGVTEJDfq5WS9wOx3gqEg9zLU&sig=AHIEtbRUS_IZGOgsJyCWUL521oc_Bo5jFg> Acesso em: 18 out 2011.

SANTOS, Ivairton M. **Data Warehouse como ferramenta de auxílio em sistemas de monitoramento ambiental**. 2003. 40 f. Monografia - Instituto de Ciências Exatas e da Terra, Departamento de Ciência da Computação Universidade Federal de Mato Grosso, 2003. Disponível em: <<http://knol.google.com/k/guilherme-mota/construindo-data-warehouse-com-software/1om0kyxjcvj20/1#>> Acesso em: 03 jun 2011.

SINGH, H. S. **Data warehouse: conceitos, tecnologias, implementação e gerenciamento**. São Paulo: Makron, 2001.

SQL Server Analysis Services. **Dados Multidimensionais**. Disponível em <<http://msdn.microsoft.com/pt-br/library/bb522607.aspx>> Acesso em: 22 set 2011.

TEOREY, Toby; LIGHTSTONE, Sam; NADEAU, Tom. **Projeto e Modelagem de Banco de Dados**. Rio de Janeiro: Editora Campus, 2007.

VALENTIM, Marta L.P. **Gestão da informação e gestão do conhecimento**. Marília, 2004. Artigo, Info Home. Disponível em: <http://www.ofaj.com.br/colunas_conteudo.php?cod=88>. Acesso em: 15 out 2011.

APÊNDICE A - Instalação e Configuração do *IREPORT*

Após fazer o download do iReports 3.7.3, inicie a instalação. Será apresentada uma tela como a mostrada na Figura 49.



Figura 49- Instalação do *iReport*
Fonte: Adaptado do *iReport*.

Clicando em *next*, será direcionado para a tela do contrato da licença. É necessário aceitar para continuar a instalação. Após aceitar os termos de uso, deixa-se o *check iReport* marcado e clica em *passa* para a próxima página, como mostra na Figura 50.

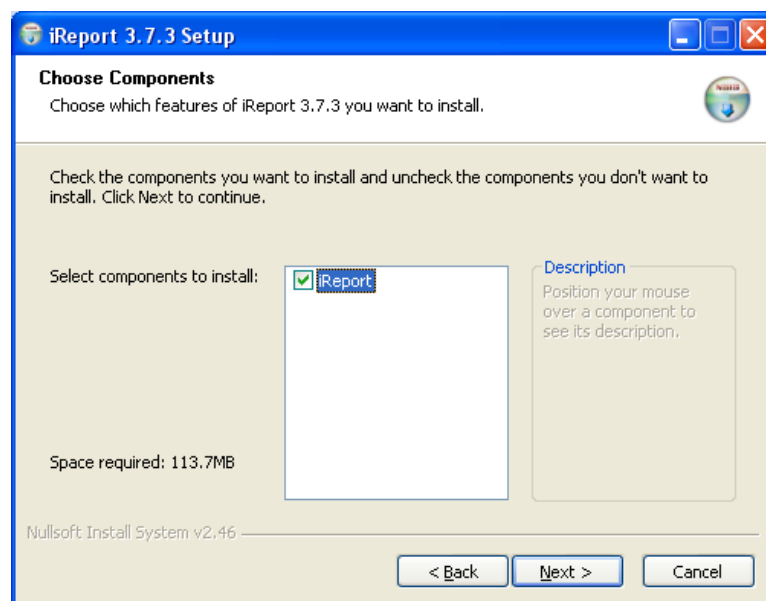


Figura 50 – Instalação do *iReport*
Fonte: Adaptado do *iReport*.

Nesse momento, faz-se necessário escolher o diretório onde será instalado a ferramenta *open source* *iReport*. Assim como na Figura 51, escolhe-se o local e prossegue-se a instalação, confirma-se o local escolhido clicando em *Install* como na Figura 52.

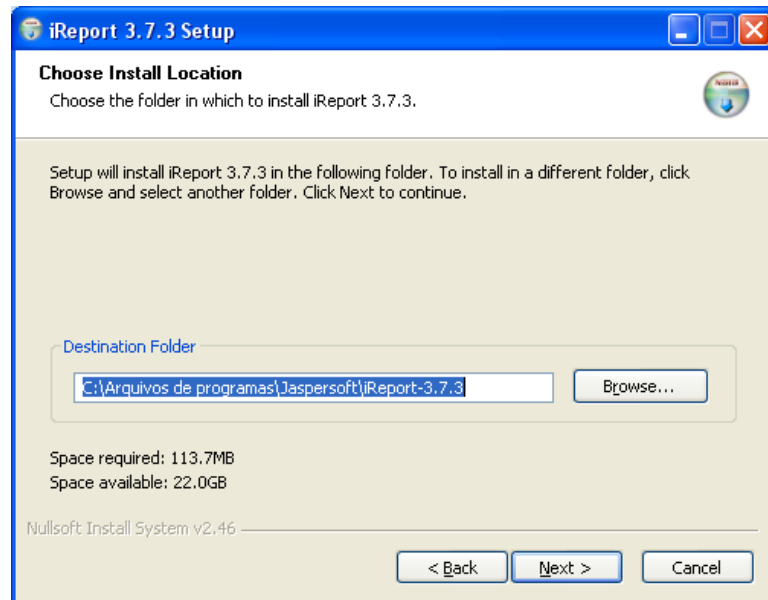


Figura 51- Instalação do *iReport*
Fonte: Adaptado do *iReport*.

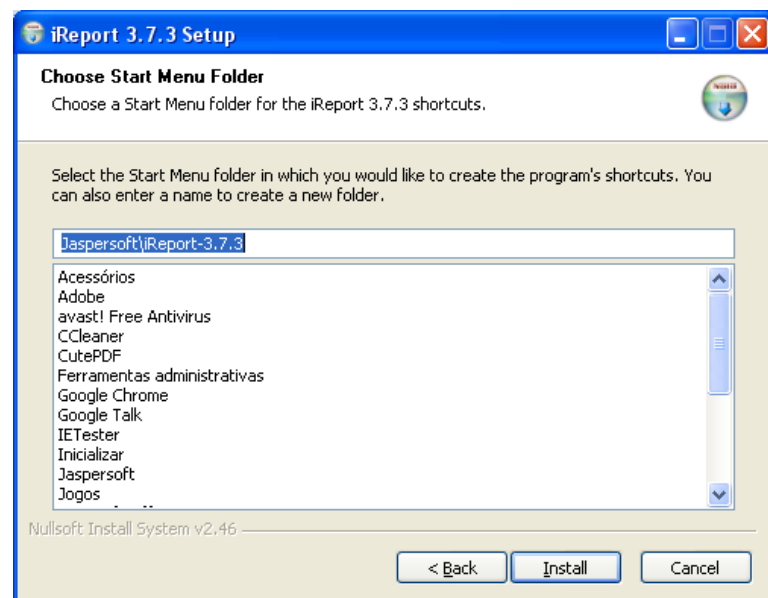


Figura 52 - Instalação do *iReport*
Fonte: Adaptado do *iReport*.

É necessário aguardar a finalização da barra de *status* como mostrado na Figura 53, para concluir a instalação.

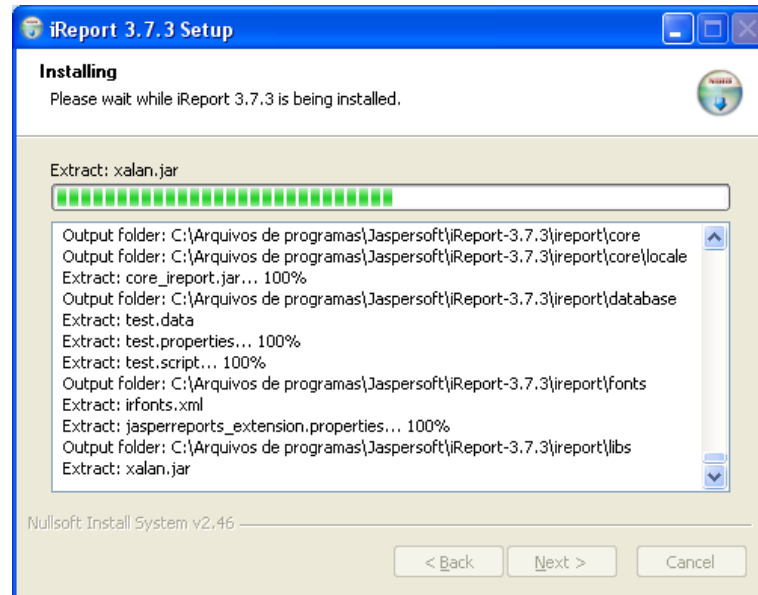


Figura 53 - Instalando o *iReport*
Fonte: Adaptado do *iReport*.

Finalizada a instalação, aparecerá uma tela solicitando a confirmação do término do instalador e para executar o *iReports*, como mostra a Figura 54.

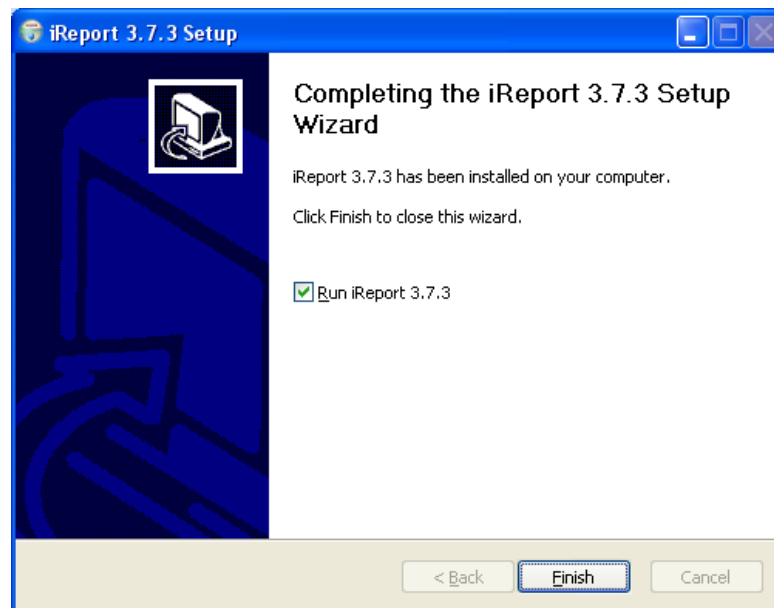


Figura 54 - Processo de Instalação Completo
Fonte: Adaptado do *iReport*.

Ao executar o aplicativo, aparecerá o ambiente de trabalho do *iReport*, como na Figura 55.

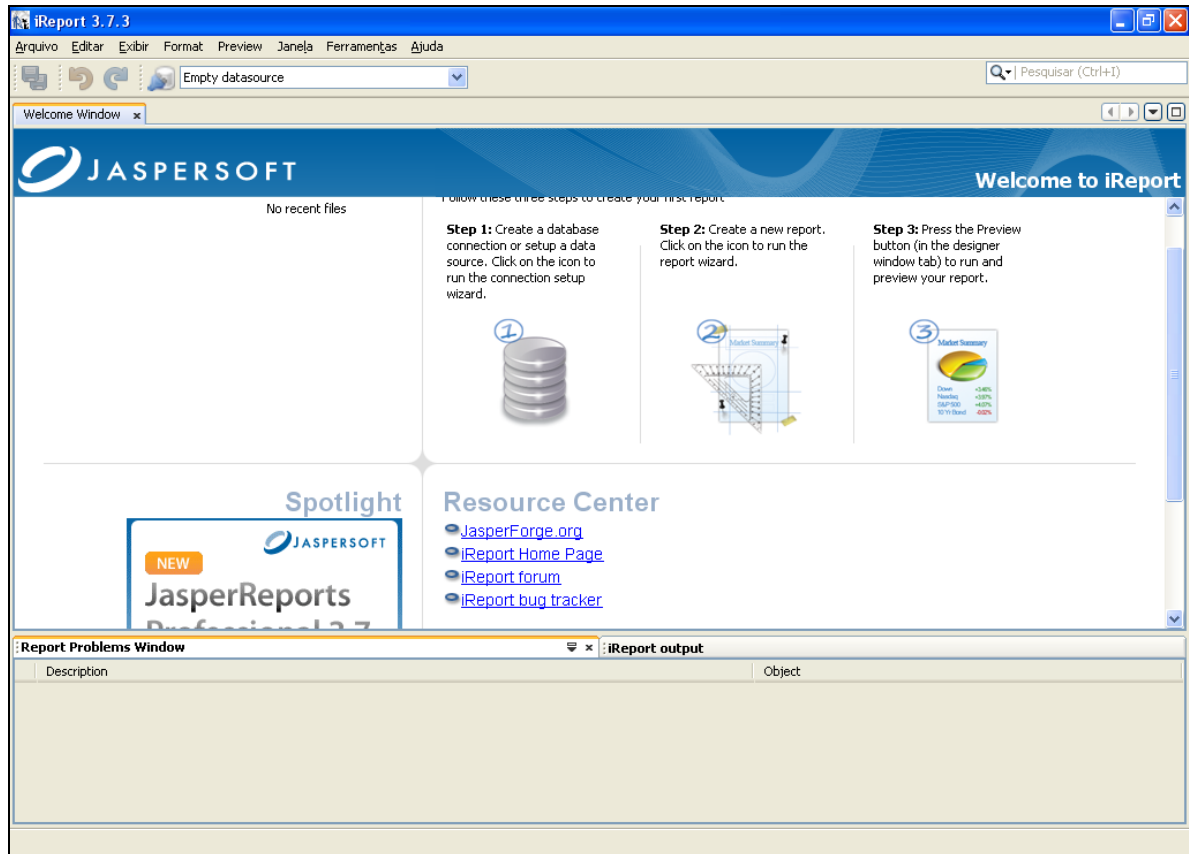


Figura 55 - Tela Inicial do iReport
Fonte: Adaptado do iReport.

Por ser a primeira vez que o *iReport* é executado, é necessário criar uma conexão com o banco de dados, clicando no ícone *Report Datasources*, como mostra a Figura 56.

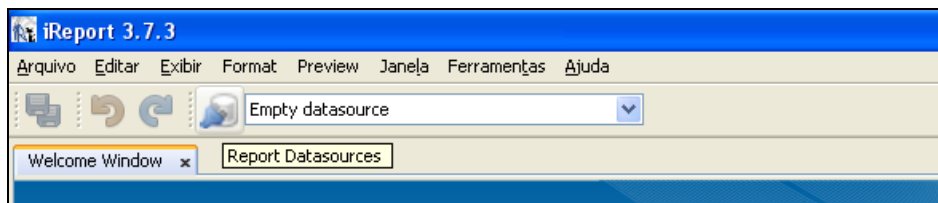


Figura 56 - Report Datasources
Fonte: Adaptado do iReport.

Será criada uma nova conexão, para isso clica-se no botão *New*, conforme a Figura 57 e será apresentada a seguinte tela, representada pelo Figura 58:

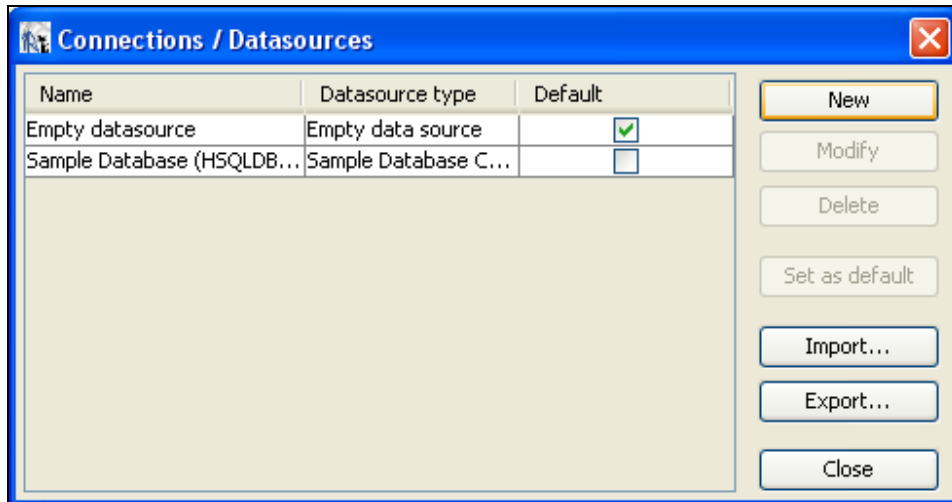


Figura 57 - Criação da nova conexão
Fonte: Adaptado do *iReport*.

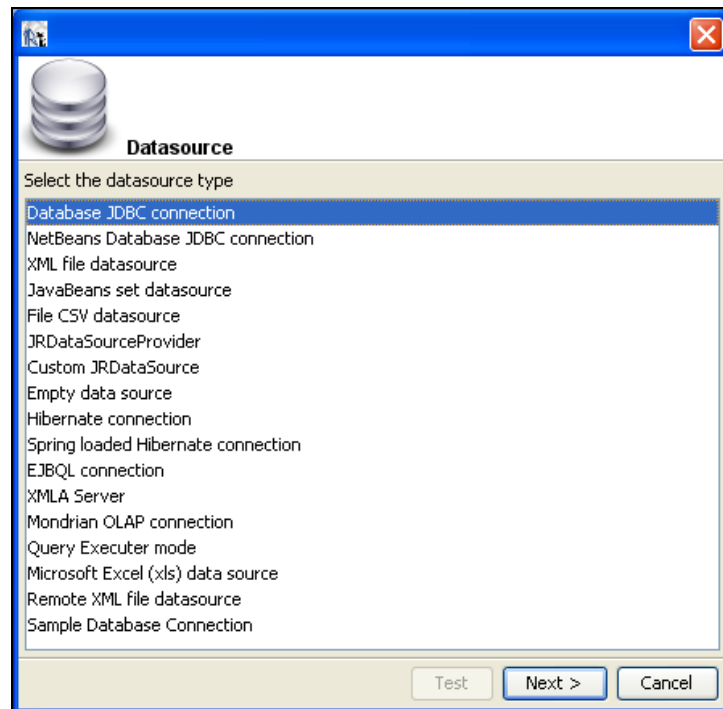


Figura 58 - Escolha do tipo de Datasource
Fonte: Adaptado do *iReport*.

Após isso, seleciona-se *Database JDBC connection*, deve-se clicar no botão prosseguir ou *next*. Como está demonstrado na Figura 59, na tela que aparecerá, falta apenas informar o nome da base de dados e, se necessário, os demais parâmetros vazios.

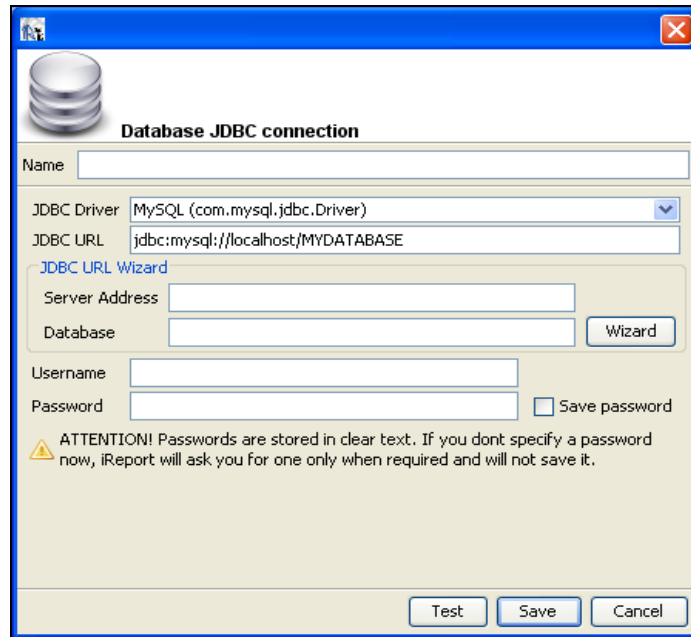


Figura 59 - Conexão com o database connection
Fonte: Adaptado do *iReport*.

Clica-se em *Test* para verificar se a conexão está correta e é válida. Aparecendo a mensagem de sucesso, como a Figura 60, clica-se em *Save*.

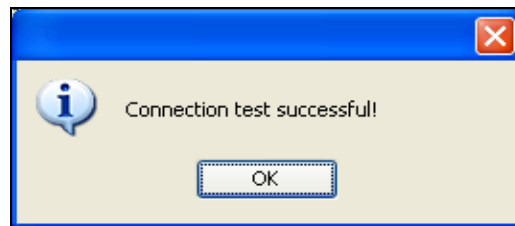


Figura 60 - Conexão com Sucesso.
Fonte: Adaptado do *iReport*.

A tela inicial de conexões será apresentada. Caso não necessite de novas conexões, poderá ser fechada.

ANEXO A - MODELO DE ENTIDADE E RELACIONAMENTO

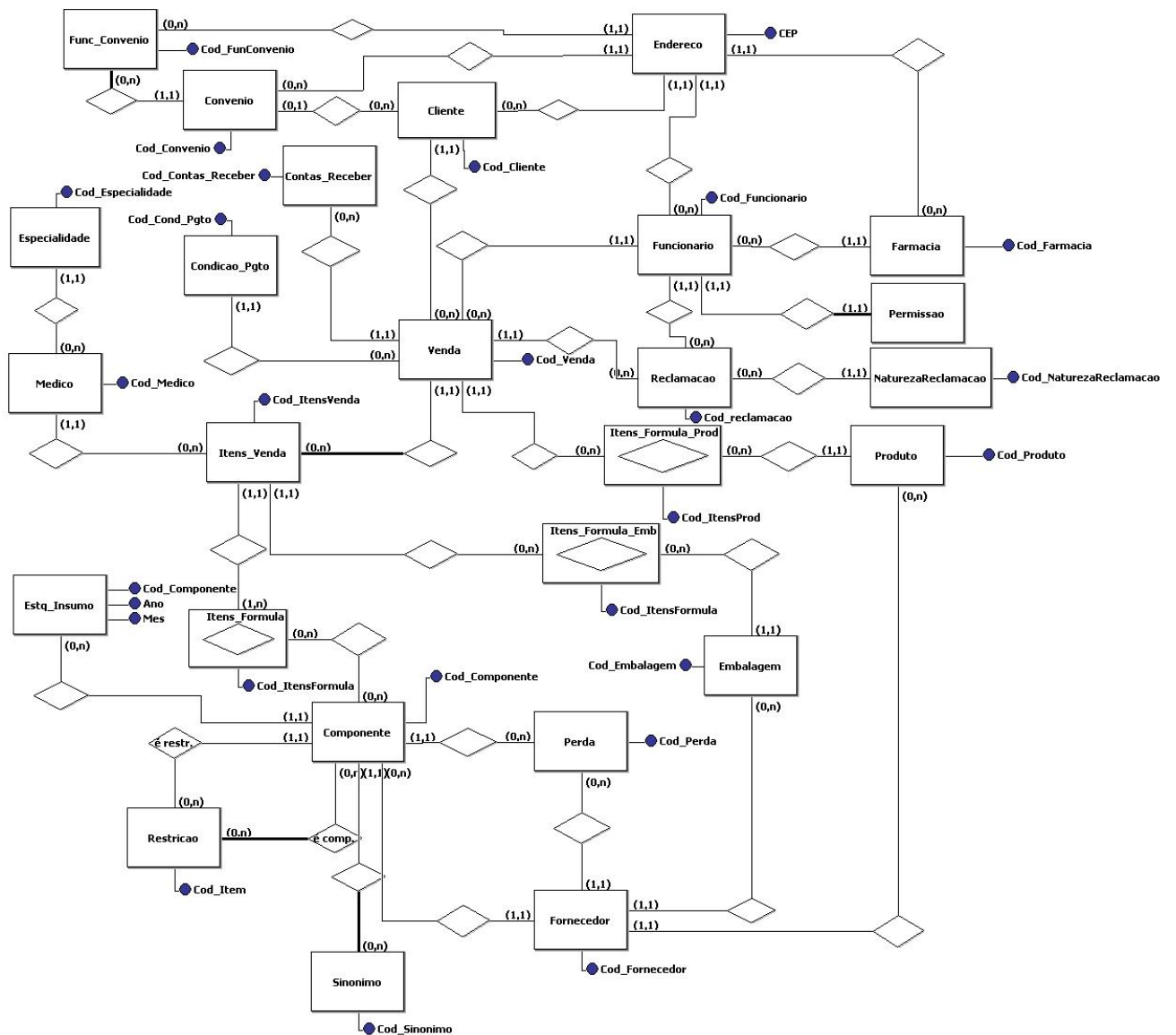


Figura 61 – Modelo de entidade e relacionamento
Fonte: IonsFarma Ltda ME

ANEXO B – DECLARAÇÃO DE CIENCIA**Ionsfarma Ltda****Rua: Gal Carneiro, 547, loja 01 e 02 – Ponta Grossa, PR****Fone/Fax: (042) 3223-2069 e (042) 3222-7893****Site: www.ionsfarma.amawebs.com****e-mail: ionsfarma@ig.com.br****DECLARAÇÃO**

Declaro para os devidos fins que conheço na íntegra o trabalho de conclusão de curso ensaiado nesta empresa e que permito, sem restrições, a divulgação do nome e demais informações fornecidas nesta empresa.

Sendo só, subscrevo-me.

Ponta Grossa, 28 de outubro de 2011

Solange de Almeida
Responsável Técnica
Proprietária

02.808.274/0001-37

IONSFARMA LTDA.

RUA GENERAL CARNEIRO, 547

84.010-370 - Ponta Grossa - PR