

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
DEPARTAMENTO ACADÊMICO DE INFORMÁTICA
CURSO DE ESPECIALIZAÇÃO EM TECNOLOGIA JAVA

DATA WAREHOUSE COM FERRAMENTAS OPEN SOURCE

Por:
FERNANDO JOSÉ ESMANIOTTO
Monografia da Especialização em Tecnologia Java

CURITIBA
2011

FERNANDO JOSÉ ESMANIOTTO

DATA WAREHOUSE COM FERRAMENTAS OPEN SOURCE

Monografia de Conclusão de Especialização, apresentado à disciplina de Trabalho de Diplomação, do curso de Especialização em Tecnologia Java do Departamento Acadêmico de Informática – DAINF – da Universidade Tecnológica Federal do Paraná – UTFPR, como requisito parcial para obtenção do título de Pós-Graduado.

Orientador: Gustavo A. Giménez Lugo

CURITIBA
2011

Em memória de Ady de Almeida Smaniotto, pessoa fantástica que nos momentos que passamos juntos sempre me incentivou a buscar meus sonhos. Aos meus pais que sempre estiveram ao meu lado. Aos meus amigos que sempre me ajudaram. Ao professor e orientador Gustavo A. Giménez Lugo que forneceu o embasamento para construção deste projeto. E em especial a DEUS, pai eterno, que iluminou minha mente e meus passos.

AGRADECIMENTOS

Certamente os poucos parágrafos que irei escrever não irão atender a as pessoas que fizeram parte desta importante fase de minha vida, portanto peço desculpas às pessoas que não estão presentes nestas palavras, mas saibam que estão guardadas em meu coração.

Ao Professor Gustavo A. Giménez Lugo obrigado pelo apoio concedido em momentos de dúvidas onde o Sr me estendeu a mão oferecendo ajuda para continuar minha caminhada.

Ao Antonio Maciel Machado através do seu apoio consegui obter sucesso em meu projeto ofereço meus agradecimentos.

Aos amigos do CRAS (Centro de Referência de Assistência Social de Mandirituba-PR), que me deram apoio para obtenção de informações cruciais para o desenvolvimento do projeto.

A Copel e Senepar que concederam dados importantes sobre o município de Mandirituba-PR para o desenvolvimento do projeto obrigado a todos.

Aos meus Pais Francisco Jose Esmaniotto e Neusa Esmaniotto que a todo tempo me incentivaram dando forças para continuar, tive diversos momentos difíceis, mas com o apoio de vocês superei todos.

A todos os professores da Universidade Tecnológica do Paraná que tive o prazer de conhecer e a todos os amigos que não citei o nome saibam que guardo dentro de meu coração as lembranças e meu eterno agradecimento.

"Habilidade é o que você é capaz de fazer. Motivação determina o que você faz. Atitude determina a qualidade do que você faz." Lou Holtz.

RESUMO

ESMANIOTTO, Fernando. Implementação de um *Data Warehouse* com ferramentas *open source* – Especialização em Tecnologia *Java*, Universidade Tecnológica do Paraná. Curitiba 2011.

Este projeto apresenta a integração de informações através de *Data Warehouse* utilizando tecnologia *Java*, os dados focalizam Mandirituba (PR) de forma que a oferecer dados específicos sobre a estrutura do município em questão, oferecem informações ricas em conteúdo para diversas áreas e profissionais, pretendendo disseminar dados para relatórios estatísticos, consultas e pesquisas. Com isso tem-se uma forma de mostrar aos gestores pontos estratégicos, onde se possam realizar ações para impedir possíveis problemas ou promover avanços.

Palavras Chave: *Data Warehouse*. Município de Mandirituba-PR. Relatórios Estatísticos. Tecnologia *Java*.

ABSTRACT

ESMANIOTTO, Fernando. Implementing a Data Warehouse with open source tools - Specialized in Java Technology, Technological University of Parana. Curitiba 2011.

This project presents the integration of information through Data Warehouse using Java technology, the data focus Mandirituba (PR) so that to provide specific data on the structure of the municipality in question, offer information-rich content for various areas and professionals, aiming to spread statistical data for reports, queries and searches. This has been a way to show managers at strategic points, where they can take actions to prevent potential problems or advancing.

Keywords: Data Warehouse. City of Mandirituba-PR. Statistical Reports. Java technology.

SUMÁRIO

CAPÍTULO 1 INTRODUÇÃO	1
1.2 OBJETIVO	2
1.3 ESTRUTURA.....	2
CAPÍTULO 2 DATA WAREHOUSE	3
2.1. ETAPAS DA IMPLEMENTAÇÃO DO DATA WAREHOUSE	3
2.2. CONSIDERAÇÕES FINAIS	7
CAPÍTULO 3 FERRAMENTAS PARA CONSTRUÇÃO DE DATA WAREHOUSE .	8
3.1 PENTAHO	8
3.1.1 Kettle	10
3.1.2 Mondrian.....	11
3.1.3 Schema Workbeach.....	12
3.1.4 Pentaho Report Designer.....	13
3.2 JASPERETL	15
3.3 TALEND OPEN STUDIO	16
3.4 CONSIDERAÇÕES FINAIS	18
CAPÍTULO 4 ESTUDO DE CASO EXTRAÇÃO, COLETA DE DADOS COM ANDROID E RELATÓRIOS	19
4.1 FONTE DOS DADOS	20
4.2 EXTRAÇÃO	21
4.2.1 Kettle	21
4.2.2 Jasperetl	23
4.2.3 Talend Open Studio	24
4.3 COLETOR DE DADOS ANDROID.....	25
4.4 RELATÓRIOS.....	28
4.5 CONSIDERAÇÕES FINAIS	29
CAPÍTULO 5 RESULTADOS	30
5.1 RELATÓRIOS.....	30
5.2 GRÁFICOS.....	32
5.3 DASHBOARD	35
5.4 CONSIDERAÇÕES FINAIS	37

CAPÍTULO 6 CONCLUSÃO E TRABALHOS FUTUROS	38
6.1 TRABALHOS FUTUROS	39
REFERÊNCIAS BIBLIOGRAFICAS	40
APÊNDICE A – Instalação da ferramenta Kettle	42
APÊNDICE B – Instalação da ferramenta Talend Open Studio.....	44
APÊNDICE C – Instalação da ferramenta JasperETL.....	48
APÊNDICE D – Instalação do servidor Mondrian	49
APÊNDICE E – Instalação da ferramenta Schema Workbench.....	52
APÊNDICE F – Instalação da ferramenta Pentaho Report Designer.....	55
APÊNDICE G – Realizando a extração de dados com a ferramenta Kettle	56
APÊNDICE H – Realizando a extração de dados com a ferramenta Talend Open Studio	62
APÊNDICE I – Realizando a extração de dados com a ferramenta JasperETL	70
APÊNDICE J – Implementando cubos com a ferramenta Schema Workbench.....	77
APÊNDICE K – Implementando relatórios com a ferramenta Pentaho Report Designer	82

LISTA DE ILUSTRAÇÕES

Figura 1 – Arquitetura do Pentaho.....	9
Figura 2 – Kettle executando uma Transformação	11
Figura 3 - Servidor Mondrian.....	12
Figura 4 - Tela Principal do Schema Workbench	13
Figura 5 - Tela Principal do Pentaho Report Designer.....	14
Figura 6 – Tela principal do JasperETL.....	15
Figura 7 - Tela Principal do Talend Open Studio.....	16
Figura 8 - Arquitetura proposta para o projeto Data Warehouse de Mandirituba (PR).....	19
Figura 9 – Mapeamento de dados.....	20
Figura 10 - Estrutura ETL realizado no Kettle.....	22
Figura 11 - Estrutura ETL realizada com o JasperETL.....	23
Figura 12 - Estrutura ETL realizada com Talend Open Studio	25
Figura 13 – Interface do coletor de dados sociais.....	26
Figura 14 – Parte código de envio de dados.....	27
Figura 15 – Parte código do <i>Web Services</i>	27
Figura 16 - Relatório e gráfico gerado pelo Pentaho Report Designer.....	28
Figura 17 - Gráfico gerado pelo <i>Pentaho Report Designer</i>	29
Figura 18 – Relatório da frota em formato: PDF, Excel e HTML	30
Figura 19 - Relatório do consumo de energia elétrica residencial	31
Figura 20 - Relatório da frota de veículos.....	31
Figura 21 – Relatório sobre empregos.....	32
Figura 22 - Gráfico sobre a frota de veículos.....	33
Figura 23 - Gráfico de empregos	33
Figura 24 - Gráfico do consumo de energia elétrica dos últimos 10 anos em Mandirituba (PR).....	34
Figura 25 – Visualização de gráfico em dispositivo Android.....	35
Figura 26 – <i>Dashboard</i> sobre o consumo de energia elétrica	36
Figura 27 - Indicador do desempenho de empregos	36
Figura 28 – Diretório de instalação do Kettle	42
Figura 29 - Tela principal do Kettle	43
Figura 30 – Tela para escolha do diretório de instalação.....	44
Figura 31 – Instalação sendo executada	44
Figura 33 – Tela contrato de licença de uso	45
Figura 33 – Cadastro na comunidade do Talend Open Source	45
Figura 34 – Escolha do repositório.....	45
Figura 35 – Configuração do repositório	46
Figura 36 – Definição do projeto.....	46
Figura 37 – Dados do projeto.....	47
Figura 38 – Tela principal do Talend Open Studio.....	47
Figura 39 – Configuração de arquivos de demonstração.....	48
Figura 40 – Diretório onde se localiza o arquivo Mondrian.war.....	49
Figura 41 – Diretório webapps do Tomcat	50
Figura 42 – Arquivo startup de inicialização do Tomcat.....	50
Figura 43 – Tela principal do servidor Mondrian.....	51
Figura 44 – Diretório de instalação do Schema Workbench.....	52

Figura 45 – Tela principal do Schema Workbench	52
Figura 46 – Configuração da conexão com a base de dados.....	53
Figura 47 – Configuração da conexão com a base de dados.....	53
Figura 48 – Executando o arquivo Launcher	55
Figura 49 – Tela principal do Pentaho Report Designer	55
Figura 50 – Tela principal do Kettle	56
Figura 51 – Tela de propriedades da Transformação	57
Figura 52 – Área de trabalho Kettle	57
Figura 53 – Tela de Leitura de Tabela	58
Figura 54 – Tela de conexão com a base.....	58
Figura 55 – Tela de Confirmação de conexão	59
Figura 56 – Tela de leitura de tabela.....	59
Figura 57 - Tela área de trabalho do Kettle	60
Figura 58 – Tela de configuração do arquivo de saída do Excel.....	60
Figura 59 – Ligação entre a base de dados e arquivo Excel através do Hop.....	61
Figura 60 – Implementação do Job.....	62
Figura 61 – Tela novo Job.....	63
Figura 62 – Área de trabalho do Talend Open Studio.....	63
Figura 63 – Inserção do componente <i>tMySqlInpu_1</i>	64
Figura 64 – Opção para realizar a conexão com a base de dados.....	64
Figura 65 – Tela de configuração da base de dados.....	65
Figura 66 – Descrição das informações sobre a base de dados.....	66
Figura 67 – Opção para realizar a configuração da consulta.....	67
Figura 68 – Configuração da Consulta.....	67
Figura 69 – Inserção do componente <i>tFileOutputExcel_1</i>	68
Figura 70 – Configuração do componente <i>tFileOutputExcel_1</i>	68
Figura 71 – Configuração da ligação entre base e arquivo de saída	69
Figura 72 – Implementação do Job com JasperETL.....	70
Figura 73 – Tela novo Job	71
Figura 74 – Job criado	71
Figura 75 – Inserção do componente <i>tMssqlInput_1</i>	72
Figura 76 – Opção para realizar configuração com banco de dados.....	72
Figura 77 – Configuração do banco de dados	73
Figura 78 – Configuração das informações sobre o banco de dados.....	74
Figura 79 – Botão para configura Consulta	74
Figura 80 – Opção para realizar a configuração da consulta.....	75
Figura 81 – Configuração da Consulta.....	75
Figura 82 – Configuração do componente <i>tFileOutputExcel_1</i>	76
Figura 83 – Ligação entra o banco de dados e o arquivo de saída	76
Figura 84 – Implementação do Schema.	77
Figura 85 – Implementação do cubo.....	77
Figura 86 – Escolha da tabela.....	78
Figura 87 – Definição de parâmetros.....	78
Figura 88 – Adicionando dimensões e hierarquias.....	79
Figura 89 – Elementos de um cubo.....	79
Figura 90 – Descrição do componente nível.....	80
Figura 91 – Descrição do componente table.....	80

Figura 92 – Configuração da senha.....	81
Figura 93 – Definição de acesso para publicação.....	81
Figura 94 – Tela de publicação do schema.....	81
Figura 95 – Tela para escolha de layout para o relatório.....	82
Figura 96 – Tela para escolha de conexão.....	82
Figura 97 – Tela de exclusão, edição ou alteração de conexão.....	83
Figura 98 – Tela para realizar a configuração da conexão.....	83
Figura 99 – Tela de inserção de comando para consulta.....	84
Figura 100 – Seleção dos dados.....	84
Figura 101 – Visualização do relatório.....	85
Figura 102 – Modificando o relatório.....	85
Figura 103 – Visualização do relatório alterado.....	86

LISTA DE TABELAS

Tabela 1 – Recursos oferecidos pela suíte Pentaho e Jasper.....	18
---	----

LISTA DE SIGLAS

ERP	Enterprise Resource Planning
ETL	Extração Transformação Carga.
OLAP	Processo analítico online
SOAP	Simple Object Access Protocol
DTS	<i>Data Transformation Service</i>
DMD	<i>Dynamic Meeting Designer</i>
JAD	<i>Join Application Developmen</i>
XML	<i>Extensible Markup Language</i>
SAD	Sistemas de apoio a decisão
BI	Business Intelligence
WEB	World Wide Web
MDX	Multidimensional Expression
SGBD	Sistema de Gerenciamento de Banco de Dados
SQL	Linguagem de Consulta Estruturada
DW	<i>Data Warehouse</i>

CAPÍTULO 1 INTRODUÇÃO

A necessidade de obter resultados rapidamente está cada vez maior. Informações que possibilitem ao gestor tomar decisões eficientes tornam-se um diferencial competitivo. O *Data Warehouse* ou Armazém de dados (DW)¹ é um grande repositório de dados, cujo principal objetivo é o de oferecer informações. “A capacidade para agir rapidamente e decisivamente num mercado cada vez mais competitivo passou a ser um fato crítico para o sucesso” (TAKAOKO, 1999). Porém, para agir com rapidez é necessário ter informações que permitam formular uma estratégia bem definida. Tendo uma estrutura organizada no DW é possível gerar informações com os quais o gestor pode formular estratégias para tomada de decisão. É baseado nesse escopo que o presente projeto pretende implementar um projeto de DW sobre o município de Mandirituba (PR).

As ferramentas utilizadas no presente projeto têm em sua base tecnologia Java. A suíte Pentaho se destaca por oferecer um conjunto de aplicativos para implementação, manipulação e apresentação de informações através de gráficos e relatórios estratégicos.

A mobilidade está cada vez mais presente no dia-a-dia das pessoas. A implementação de um recurso que possibilite resgatar informações a qualquer momento o torna um diferencial competitivo. O Projeto aborda a implementação de um mecanismo para coleta e apresentação de informações utilizando a tecnologia Android conectado a *Web Services*.

Android é um sistema operacional móvel de código aberto o qual permite desenvolver software em linguagem Java. *Web Services* é uma tecnologia que traz mobilidade e independência através dela é possível integrar qualquer tipo de tecnologia.

Ao longo do projeto serão apresentados conceitos, ferramentas, estudos de caso sobre o município, manuais de instalação e os resultados obtidos com a utilização das ferramentas da suíte Pentaho. Permitindo ao leitor obter informações sobre as ferramentas *open source* (software livre) para implementação do DW.

¹ Durante o presente projeto o termo Data Warehouse será abreviado pela sigla DW.

1.2 OBJETIVO

O objetivo do presente projeto é de demonstrar a utilização de software *open source* (software livre) para criação de um projeto de DW. O foco principal se dá em torno das ferramentas da suíte Pentaho, as quais utilizam em sua base tecnologia Java. O contexto no qual o ambiente do DW será implementado está direcionado ao município de Mandirituba (PR). Através de relatórios e gráficos será possível observar aspectos importantes de crescimento do município. A solução da criação de um projeto de DW aqui apresentada tem o propósito de atuar como infra estrutura para o fornecimento de informações de qualidade.

1.3 ESTRUTURA

O presente projeto está organizado da seguinte forma: no capítulo 2 é abordados os conceitos sobre a implementação de um projeto de DW. O Capítulo 3 apresenta conceitos das ferramentas que serão utilizadas. O capítulo 4 descreve aspectos da estrutura das informações obtidas sobre o município e também sobre o processo de extração, transformação e carga (ETL)². Visualização de informações em dispositivos móveis e a apresentação de relatórios e gráficos. O capítulo 5 apresenta os principais resultados obtidos com a utilização das ferramentas da suíte Pentaho. Por fim, no capítulo 6 são descritas as conclusões da monografia, as contribuições e trabalhos futuros.

² Durante o presente projeto o termo extração, transformação e carga será abreviado pela sigla ETL.

CAPÍTULO 2 DATA WAREHOUSE

Informações são essenciais para tomada de decisão. Ter informações válidas, confiáveis e de qualidade oferece ao gestor uma maneira para formular estratégias bem definidas. O DW se baseia na construção de um repositório de informações específicas das atividades de uma organização com sua utilização é possível obter relatórios, gráficos e informações válidas para a tomada de decisão.

“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; trata-se de um processo em andamento que aglutina dados de fontes heterogêneas, incluindo dados históricos.” (INMON, 2005).

O DW oferece recursos para análise de grandes quantidades de informações. “Um banco de dados é uma coleção de dados operacionais armazenados e utilizados pelo sistema de aplicações de uma empresa específica. Os dados mantidos por uma empresa são chamados de ‘operacionais’ ou ‘primitivos’” (BATINI; LENZERINI, 1986). Dados operacionais fazem referência a banco de dados transacionais. Porém, existem diversas fontes de dados que podem ser utilizadas para popular o DW tais como: arquivos de texto, planilhas, etc. Os dados armazenados no DW não são voláteis, ou seja, não mudam, exceto quando é necessário realizar alterações ou correção de alguma informação.

2.1 ETAPAS DA IMPLEMENTAÇÃO DO DATA WAREHOUSE

Analisar a infra estrutura e definir uma abordagem para implementação de um projeto de DW é fundamental para seu sucesso. A seguir é apresentada uma relação que define uma abordagem estruturada para criação de um projeto de DW.

- 1) Armazenamento:** Armazenamento de dados em DW é feito em um único depósito, o qual deve ser de rápido acesso para futuras análises. Quanto mais dados contiver o depósito, melhor a análise.
- 2) Modelagem de Dados:** A modelagem dos dados de DW é chamada de – normalizados. Com isso tem um significativo aumento no desempenho em

consultas, relatórios além de propiciar um modelo de processo intuitivo para quem o utiliza.

3) Características: “*Data Warehouse* deve ser orientado por assuntos, integrados, variáveis no tempo e não voláteis” (INMON, 2005). Sendo estas as principais características de um DW, as quais serão descritas com maiores detalhes logo abaixo:

- **Orientação por Assunto:** Principal característica do projeto DW. A modelagem se dá em torno dos principais assuntos da empresa. Enquanto os sistemas transacionais são voltados para processos o armazém de dados é voltado para assuntos.

- **Integração:** Característica importante do projeto DW, etapa onde se perde mais tempo. Definido a grande quantidade de informações de fontes de dados diferentes.

- **Variação no Tempo:** “Os *Data Warehouses* são variáveis em relação ao tempo” (INMON, 2005). Isso quer dizer que os dados armazenados em DW, ficam por um período de tempo superior a de um sistema transacional.

No caso do DW, o principal objetivo é a análise do comportamento das informações durante um período de tempo maior. Com isso, as decisões tomadas pelos gestores se baseiam em fatos que ocorrem em determinado período. Logo então, se tem informações que vão sendo armazenadas ao longo do tempo e, cada vez que são armazenadas, se tem dados específicos do determinado período. Com isto, é possível formular estratégias organizacionais com base nos dados anteriores.

É importante ter em mente que os metadados também possuem elementos temporais, porque mantém históricos das mudanças das regras de negócio da empresa. Os metadados são elementos responsáveis pelas informações referentes ao caminho dos dados dentro do DW.

- **Não Volatilidade:** Dentro do DW são realizadas apenas duas operações: carga e consulta diferente, por exemplo, de um sistema transacional onde são realizadas operações de inserir, atualizar, selecionar, além de apagar registros.

Deve-se considerar que os dados, antes de serem armazenados no DW, passam por filtros e muitas informações nem chegam a ser inseridas no armazém de dados.

4) Localização: Os dados podem estar fisicamente armazenados de três formas:

- **Centralizado:** Em um local centralizando tornando as consultas dos dados eficientes. Porém em termos de investimento torna-se um pouco inconveniente, pois tem um grande gasto com servidores para abrigar a estrutura DW.

- **Distribuídos:** *Data Marts* é uma alternativa bastante interessante, pois, é possível armazenar dados por áreas de interesse, por exemplo, o setor de Marketing de uma empresa. Em vez de se ter uma estrutura gigantesca, tem-se uma estrutura por área específica.

- **Por níveis de detalhes:** Trata-se de uma forma de armazenamento separado onde informações. É possível ter informações sobre vendas em um servidor e informações sobre marketing em outro.

5) Credibilidade dos dados: É um instrumento importante para o sucesso de qualquer projeto. No DW, a credibilidade de dados é fundamental. Pequenas inconsistências podem causar sérios danos na hora de extrair informações para formular uma estratégia. Dados não dignos de confiança tornam consultas ou relatórios inúteis.

6) Granularidade: Granularidade é um resumo dos dados existentes em DW, ou seja, quanto maior for o nível de detalhe, menor será a granularidade. Esse processo, afeta diretamente o volume de dados armazenado no armazém de dados e ao mesmo tempo o tipo de consulta.

7) Metadados: Os metadados são referência a outros dados. As etapas da implementação de um projeto de DW, que vai desde a modelagem até a visualização das informações gera metadados. É nos metadados que ficam armazenados informações sobre, por exemplo: atributos das tabelas, descrições sobre a estrutura, histórico de mudanças etc.

Segundo (INMON, 2005), os metadados mantêm informações sobre “o que está e onde”. Os aspectos nos quais os metadados mantêm informações são:

- A estrutura dos dados, visão do programador;
- A estrutura dos dados, visão dos analistas de Sistema de apoio a decisão (SAD);
- A fonte de dados que alimenta DW;

- A transformação sofrida dos dados no momento de sua migração para DW;
- Modelo de dados;
- O relacionamento entre modelo de dados e o DW;
- O histórico das extrações de dados.

8) Processos de carga: “A extração de dados é um processo complexo devido as diversas tecnologias de armazenamento existentes na atualidade” (GONÇALVES, 2002). Envolve a fase de extração dos dados de sistemas legados e de outras fontes, tais como planilhas e arquivos de texto. Filtragens são realizadas para garantir a integridade dos dados que serão armazenados no armazém de dados. Basicamente o processo de carga de dados envolve três etapas: extração, filtragem e a carga de dados.

9) Metodologia de levantamento: É a fase da implementação do DW que por muitas vezes é esquecida pelos profissionais que programam o armazém de dados. Trata-se do levantamento de dados gerenciais que são indispensáveis para o sucesso de um bom Sistema de Apoio à Decisão (SAD).

No momento em que se fala da implementação de DW, muitos profissionais logo pensam na extração de dados de sistemas legados para dar carga ao futuro armazém de dados.

10) Processo analítico online (OLAP): “É uma ferramenta muito importante no contexto gerencial, ajudando a analisar de forma mais eficiente, a quantidade de dados crescente armazenada pelas organizações, transformando-os em informação útil” (THOMSEN, 2002).

Processamento analítico online (OLAP) tem como objetivo extrair informações do armazém de dados gerando relatórios, consultas e, portanto, servindo os gestores com informações estratégicas para auxiliar na tomada de decisão.

2.2 CONSIDERAÇÕES FINAIS

Neste capítulo foram abordados os principais conceitos para implementação de um projeto de DW. Além das etapas necessárias para criação de um projeto de estruturado. O próximo capítulo aborda conceitos das ferramentas que serão utilizadas para implementar o projeto de DW de Mandnrituba (PR).

CAPÍTULO 3 FERRAMENTAS PARA CONSTRUÇÃO DE DATA WAREHOUSE

Para implementar o projeto do DW de Mandirituba (PR), serão utilizadas ferramentas *open source* (software livre) para realizar o processo de extração, transformação e carga (ETL), implementação das análises de processo analítico online (OLAP) e também para implementação de relatórios e gráficos. Questões como usabilidade, desempenho e confiabilidade do software serão apresentados.

A monografia é voltada para demonstrar a utilização das ferramentas da suíte Pentaho que é baseada em tecnologia Java, sendo atualmente uma das mais ricas no conceito de *Business Intelligence* ou inteligência empresarial (BI)³.

Além das ferramentas da suíte Pentaho, a monografia irá abordar os conceitos de outras duas ferramentas para realizar o processo de ETL o JasperETL (JASPERSOFT, 2011) e o Talend Open Studio (TALEND, 2011), ferramentas *open source* (software livre).

3.1 PENTAHO

A Plataforma Pentaho, baseia-se na construção de soluções para BI. “Pentaho é uma suíte poderosa de Inteligência de Negócios que oferece muitos benefícios como: relatórios, tabelas dashboarding processamento OLAP” (BOUMAN; DONGEN, 2009). Desenvolvido utilizando tecnologia Java o software oferece uma arquitetura voltada para WEB e executa nos principais servidores, tais como: Tomcat (CHOPRA; LI; GENENDER, 2007) e Jboss (JAMAR; JOHNSON, 2009). “O Pentaho BI Suíte é um software de código aberto, você está livre para usar e distribuir” (BOUMAN; DONGEN, 2009). Possui estrutura integrada para soluções de BI, trata-se de um sistema unificado proporcionando redução na sobrecarga de desenvolvimento de soluções para inteligência empresarial. A figura 1 apresenta a arquitetura do Pentaho.

³ Durante o presente projeto o termo Business Intelligence será substituído pela sigla BI.

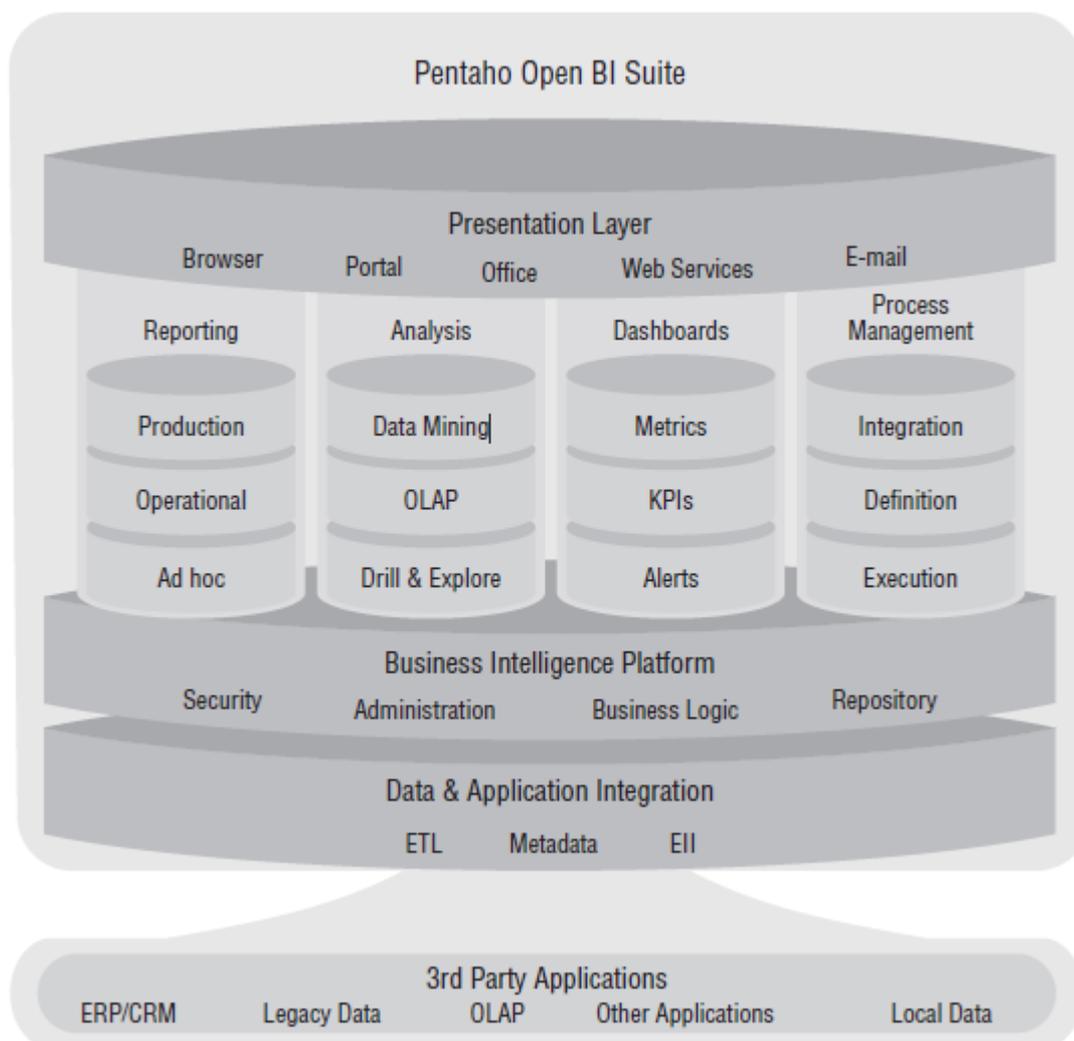


Figura 1 – Arquitetura do Pentaho
Fonte: BOUMAN; DONGEN (2009)

“Pentaho é um conjunto de inteligência de negócios ao invés de um único produto: ela é feita por um conjunto de programas de computador que trabalham juntos para criar e oferecer soluções de *Business Intelligence*” (BOUMAN; DONGEN, 2009). A figura 1 apresentou a maneira que a plataforma é integrada, onde os módulos comunicam-se entre si. A seguir apresentados os conceitos das principais ferramentas utilizadas para montar o DW de Mandirituba (PR). Ao final deste capítulo será apresentada uma tabela que apresenta os recursos que a suíte Pentaho e a suíte Jasper oferecem.

3.1.1 Kettle

A ferramenta Kettle é utilizada para realizar processo de ETL, fazendo parte da suíte Pentaho. “Arquivos são os mais primitivos, mais também o formato mais utilizado armazenamento de dados. A ferramenta Kettle tem a capacidade de ler dados de todos os tipos de arquivos e formatos diferentes” (PULVIRENTI; ROLDÁN, 2011). Frequentemente utilizada em projetos de DW a ferramenta oferece versatilidade em sua utilização. A ferramenta também pode ser usada para outros propósitos, tais como: migração de dados entre aplicações ou base de dados, exportação de informações de banco de dados para arquivos, limpeza de dados e na integração de aplicações. A seguir são listados alguns dos principais conceitos da ferramenta:

- **TRANSFORMAÇÃO:** Trata-se de uma rotina formada por uma coleção de passos interligados, onde a principio tem a representação das fontes de dados e ao final a saída dos dados.
- **Job (Trabalho):** O Job é uma rotina para execução das transformações, sendo que um Trabalho pode executar mais de uma transformação ou até mesmo outros Trabalhos.
- **Step (Passo):** É cada Step dentro de uma transformação. Através dos Steps torna-se possível organizar as transformações.
- **Hop (Salto):** É a representação gráfica do fluxo dos dados entre dois Steps, onde se tem a origem e o destino, dados que passam pelo Hops constituem os dados de saída.

A figura 2 demonstra os conceitos mencionados acima sobre a ferramenta Kettle.

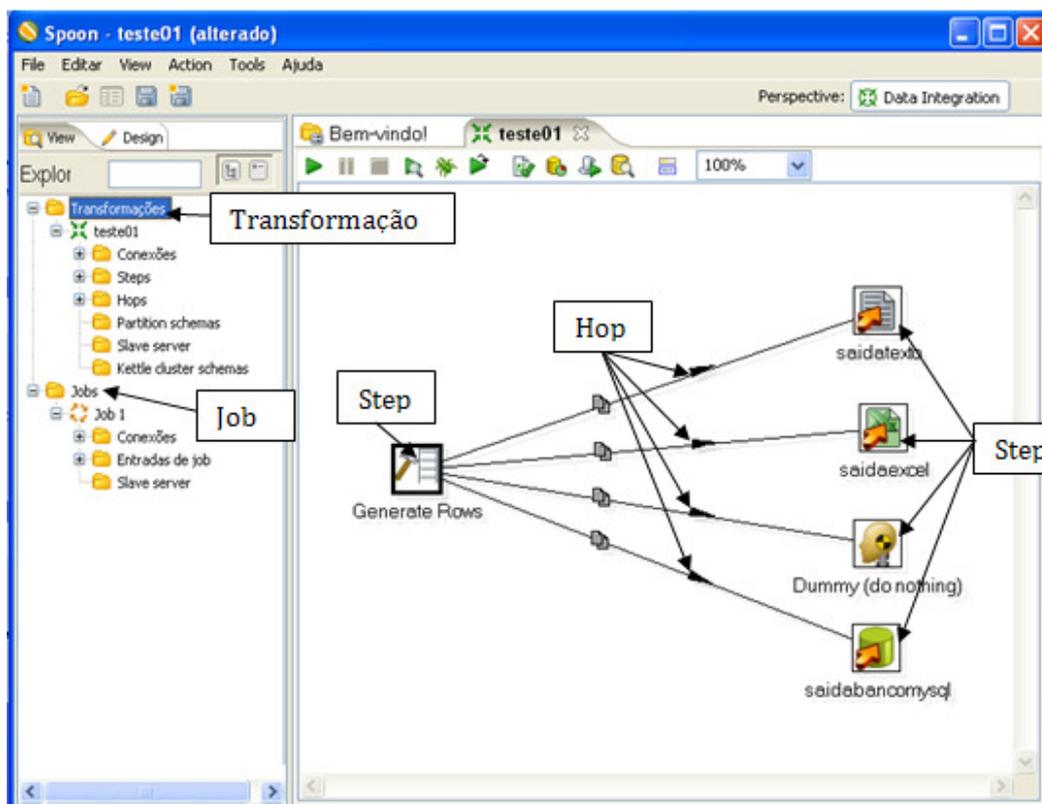


Figura 2 – Kettle executando uma Transformação
 Fonte: Autoria própria.

“Transformação é uma entidade feita de etapas ligadas por ‘Hop’” (ROLDÁN, 2010). A figura 2 mostra a execução de uma rotina de Transformação, onde o ambiente é composto por **Steps** interligados por **Hops**. O Kettle disponibiliza diversos tipos de **Steps** para realizar as mais diversas ações seja dentro de uma Transformação ou Trabalho. No Apêndice A encontra-se um manual que demonstra o processo necessário para realizar a instalação da ferramenta. Além disso, o Apêndice G demonstra com detalhes o processo de ETL usando a ferramenta Kettle.

3.1.2 Mondrian

A ferramenta Mondrian é utilizada como servidor para Processamento Analítico Online (OLAP), que significa processar várias informações em tempo real, fazendo parte da suíte Pentaho. “O núcleo do Mondrian é um JAR que atua como “JDBC para OLAP”: fornecer conexões e executar SQL contra a base de dados relacionais” (PENTAHO ANALYSIS SERVICES, 2011).

Desenvolvido utilizando tecnologia Java, o Mondrian executa consultas construídas na linguagem de expressões multidimensionais (MDX) que basicamente são comandos SQL com sua sintaxe modificada, utilizados para manipular dados multidimensionais. A figura 3 apresenta o servidor em execução.

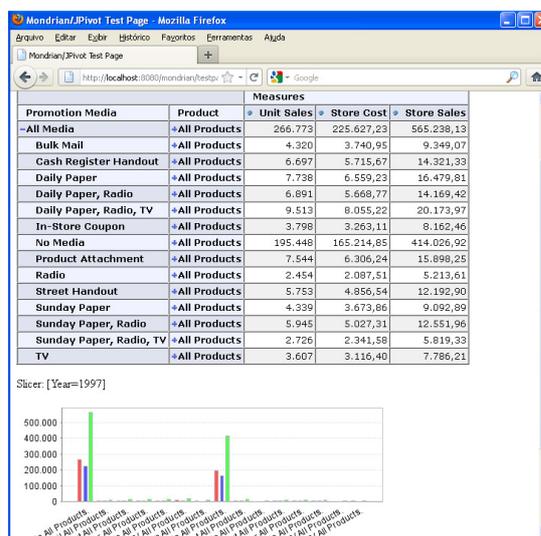


Figura 3 - Servidor Mondrian
Fonte: Autoria própria.

A figura 3 apresenta o servidor executando a análise. Note que é possível ter a visualização das informações em forma de relatório e também em forma gráfica. No Apêndice D existe um manual que demonstra o processo necessário para realizar a instalação da ferramenta.

3.1.3 Schema Workbench

“O Schema Workbench é uma interface designer que permite criar e testar esquemas cubo Mondrian OLAP visualmente” (PENTAHO ANALYSIS SERVICES, 2011). A figura 4 mostra a interface da ferramenta.

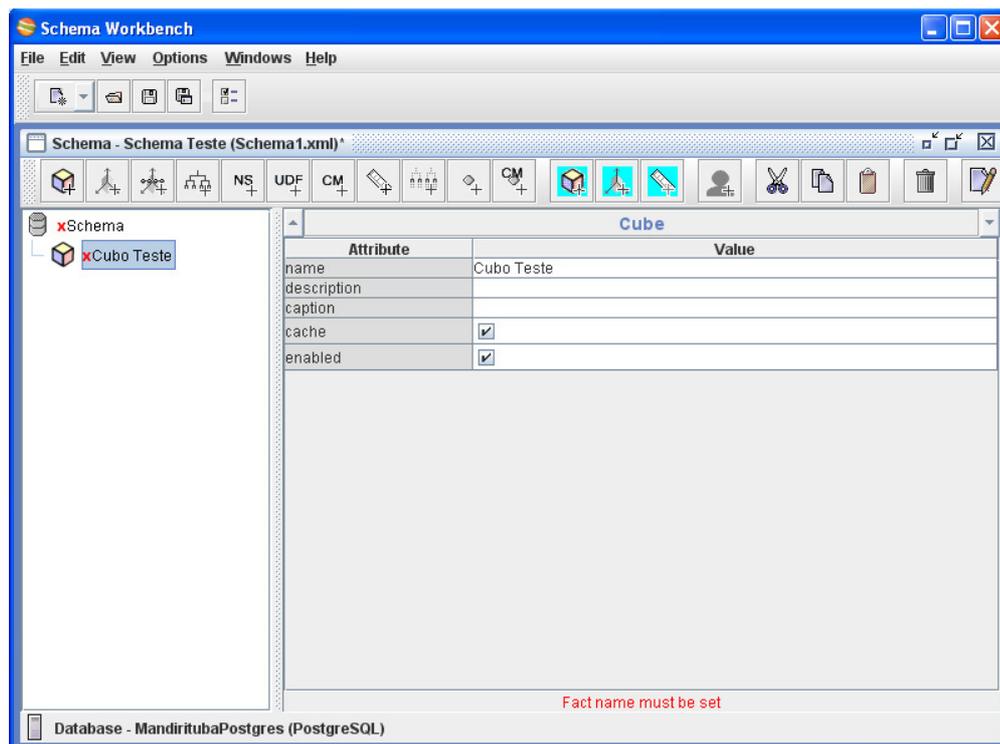


Figura 4 - Tela Principal do Schema Workbench
Fonte: Autoria própria.

“Este modelo pode ser considerado um cubo, que são estruturas que utilizam FATO existente em uma tabela DIMENSÃO encontrado em seu SGBD” (PENTAHO ANALYSIS SERVICES, 2011). Basicamente isso quer dizer que o aplicativo irá se conectar a um banco de dados relacional montando uma estrutura em forma de dimensões. Posteriormente o servidor Mondrian interpreta esta estrutura e apresenta ao usuário na forma de gráfico. No mesmo é possível interagir entre as dimensões, ou seja, a estrutura do cubo pode conter uma dimensão de tempo onde se torna possível navegar entre períodos diferentes e ao mesmo tempo observar os resultados.

Detalhes sobre os procedimentos para gerar um cubo de dados são encontrado no Apêndice J apresenta os passos para implementação e manipulação do cubo. Além disso, o Apêndice E explica o processo de instalação do aplicativo.

3.1.4 Pentaho Report Designer

A ferramenta Pentaho Report Designer faz parte da suíte Pentaho voltado para criação de relatório. “Usuários de negócios precisam de acesso informações

em muitas formas diferentes para as mais diversas razões” (GORMAN, 2009). A ferramenta oferece um mecanismo para o usuário obter informações através de relatórios e gráficos. A figura 5 mostra a interface da ferramenta e também uma demonstração de um relatório gerado.

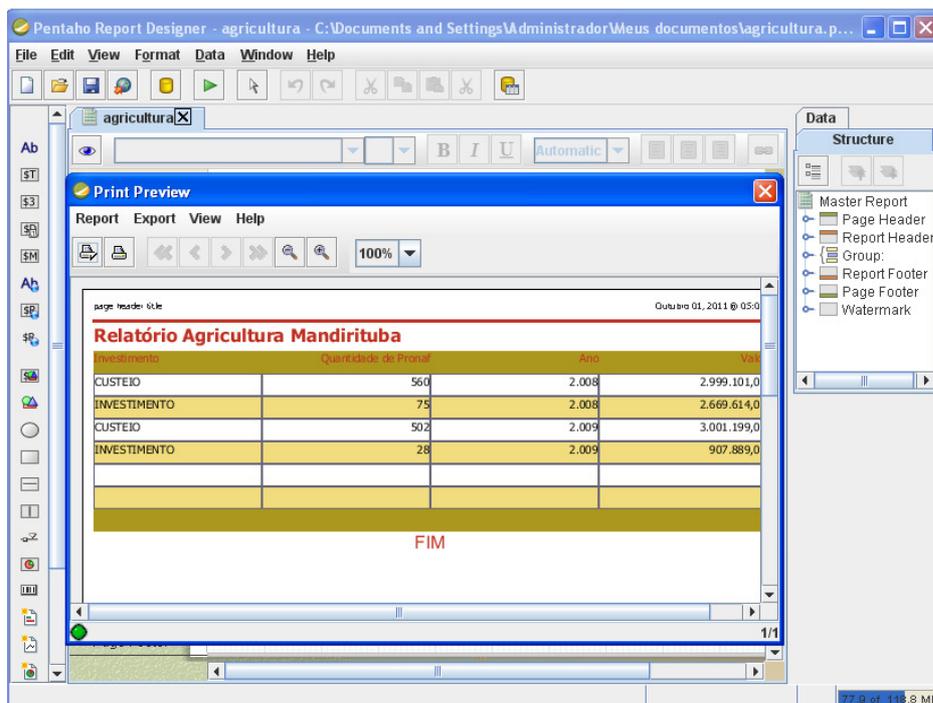


Figura 5 - Tela Principal do Pentaho Report Designer
Fonte: Autoria própria.

Basicamente o software se conecta a uma fonte de dados como, por exemplo, um banco de dados relacional. Através de um esquema de layout pré definido gera os relatórios ou gráficos. No Apêndice 6 encontra um manual de instalação da ferramenta, além disso o Apêndice K mostra com detalhes o processo para conexão com a fontes de dados e a geração de relatórios.

Até este ponto, foram demonstrados os conceitos das ferramentas Kettle, Mondrian, Schema Workbench e Pentaho Report Designer que compõem a suíte de aplicativos Pentaho as quais serão utilizadas para implementação do DW do município de Mandirituba (PR). As seções 3.2 e 3.3 abordam conceitos de outras duas ferramentas para realizar o processo de ETL. Porém essas ferramentas não fazem parte da suíte Pentaho.

3.2 JASPERETL

A ferramenta JasperETL (JASPERSOFT, 2011) é uma ferramenta *open source* (software livre) para integração e tratamento de dados. Focado no ambiente de desenvolvimento o JasperETL automatiza rotinas de integração de sistemas heterogêneos através da importação e exportação de dados.

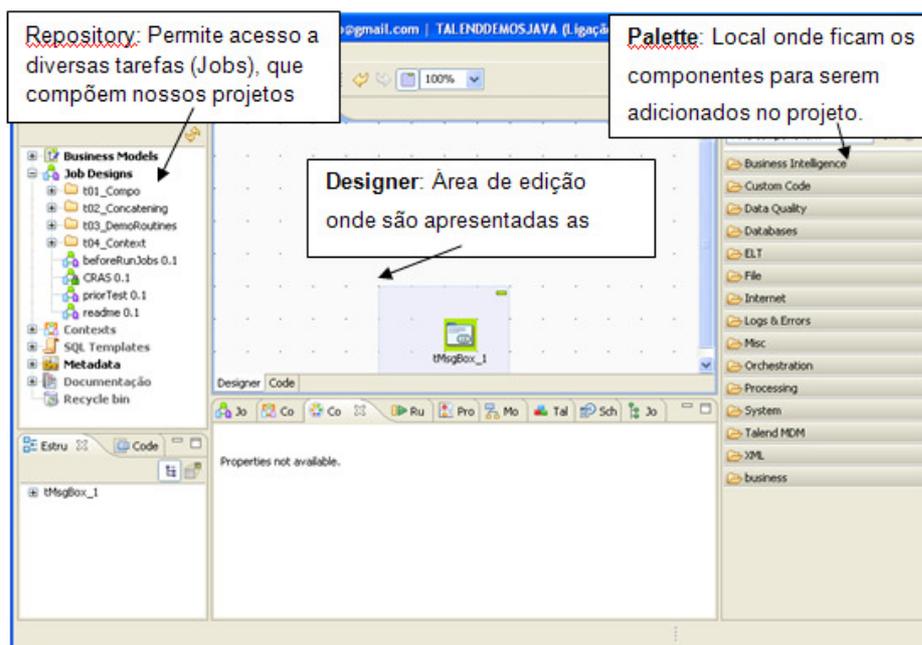


Figura 6 – Tela principal do JasperETL
Fonte: Autoria própria.

A seguir serão abordados os conceitos sobre o Repositório local onde são armazenados os elementos do projeto:

- **Business Models (Modelos de negócio):** Objetivo desta categoria é para fins de documentação. Nesta parte podem ser construídas notas explicativas que representem em uma linguagem mais acessível o objetivo do projeto. Apresenta elementos para criar um fluxograma de seqüência, de operações que representam o processo, onde também é possível anexar qualquer tipo de arquivo para documentar o projeto.
- **Job Designs (Desenho de trabalho):** Local onde ficam as tarefas, é possível também criar pastas para organizar as tarefas no projeto.
- **Contexts (Contextos):** É o local onde são definidas as constantes e seus respectivos valores agrupando em contextos.

- **SQL Templates (Modelos de SQL):** É possível definir comandos SQL para serem executados dentro do projeto.
- **Metadata (Metadados):** Categoria de grande importância, pois permite definir e configurar os recursos que as tarefas vão necessitar no decorrer do projeto, ex: conexão com banco de dados, arquivos externos, *Web Service* entre outros.
- **Documentação:** Local onde é possível armazenar qualquer tipo de arquivo para fins de documentação do projeto.
- **Recycle bin (Lixeira):** Local onde ficam os elementos que são apagados, estando na Lixeira é possível recuperar.

No Apêndice C existe um manual que demonstra o processo necessário para realizar a instalação da ferramenta. Além disso, o Apêndice I demonstra com detalhes o processo de extração, transformação e carga usando a ferramenta.

3.3 TALEND OPEN STUDIO

O Talend Open Studio (TALEND, 2011) é uma ferramenta *open source* (software livre) para realizar o processo de ETL. Através do Talend é possível realizar o procedimento de migração de base de dados e integração. A figura 7 mostra a interface e também explica alguns dos principais itens que compõem a ferramenta.

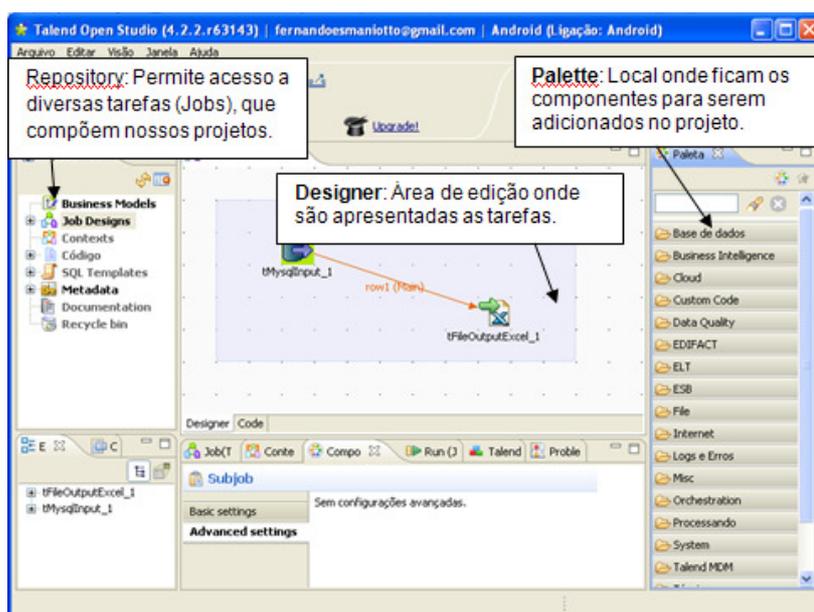


Figura 7 - Tela Principal do Talend Open Studio
 Fonte: Autoria própria.

A seguir serão explicados conceitos sobre o Repositório local onde são armazenados os elementos do projeto:

- **Busines Models (Modelos de negócio):** Usado para fins de documentação. Nesta parte podem ser construídas notas explicativas que representem em uma linguagem mais acessível o objetivo do projeto. Apresenta elementos para criar um fluxograma de sequência, de operações que representam o processo, onde também é possível anexar qualquer tipo de arquivo para documentar o projeto.
- **Job Designs (Desenho de trabalho):** Local onde ficam as tarefas. É possível também criar pastas para organizar as tarefas no projeto.
- **Contexts (Contextos):** É o local onde são definidas as constantes e seus respectivos valores agrupando em contextos.
- **SQL Templates (Modelos de SQL):** Local onde é possível definir comandos SQL para serem executados dentro do projeto.
- **Metadata (Metadados):** Categoria de grande importância permite definir e configurar os recursos que as tarefas vão necessitar no decorrer do projeto, ex: conexão com banco de dados, arquivos externos, *Web Service* entre outros.
- **Documentation (Documentação):** Local usado para armazenar qualquer tipo de arquivo para fins de documentação do projeto.
- **Recycle bin (Lixeira):** Local onde ficam os elementos que são apagados, estando na Lixeira é possível recuperar.

No Apêndice B encontra-se um manual que demonstra o processo necessário para realizar a instalação da ferramenta. Além disso, o Apêndice H demonstra o processo para realizar o ETL.

A seguir é apresentada na tabela um comparativo dos recursos oferecidos pela Plataforma Pentaho e também pela ferramenta Jasper.

Tabela 1 – Recursos oferecidos pela suíte Pentaho e Jasper

MÓDULOS	PENTAHO	JASPER
Servidor de Aplicação	JBoss	JBoss
Autenticação de usuário	Acegi	Acegi
Colaboração	JFreeChart	JFreeChart
Dashboard	-	-
Data Mining	-	Weka
DBMS	MySQL, Oracle, SQL Server, PostgreSQL, etc.	MySQL, Oracle, SQL Server, PostgreSQL, etc.
ETL	JasperETL	Pentaho Data Integration
Geo-referenciamento	Google Maps	Google Maps
Agendador de tarefas	Quartz	Quartz
OLAP	Mondrian&Jpivot	Mondrian&Jpivot
Portal	JBoss Portal	Liferay
Relatórios	Pentaho Report Designer, JasperReport, BIRT	JasperReport
Consultas de exemplo	-	-
Login único	CAS	Acegi
Servidor Web	Tomcat	Tomcat

Fonte: GOLFARELLI (2009)

A tabela 1 apresentou os recursos que a suíte de aplicativos Pentaho e Jasper oferecem para criação de um projeto de DW. Porém na monografia serão utilizados os recursos do aplicativo JasperETL, o qual utilizado para realizar o processo de ETL.

3.4 CONSIDERAÇÕES FINAIS

Este capítulo abordou os principais conceitos sobre as ferramentas da suíte Pentaho. Também foi abordado conceitos das ferramentas JasperETL e Talend Open Studio, usadas para realizar o processo de ETL. Além dos conceitos sobre as ferramentas, foi apresentada uma tabela que mostra os recursos que a suíte Pentaho e Jasper oferece. No próximo capítulo inicia-se uma série de estudos de caso demonstrando a utilização das ferramentas para realizar a construção do projeto de DW de Mandirituba (PR).

CAPÍTULO 4 ESTUDO DE CASO EXTRAÇÃO, COLETA DE DADOS COM ANDROID E RELATÓRIOS

A partir deste ponto serão apresentados estudos de caso de extração, coleta de dados com Android e relatórios baseados nas informações levantadas sobre o município de Mandirituba (PR).

Para realizar o processo de ETL serão utilizadas as ferramentas: Kettle da suíte Pentaho, JasperETL da suíte JasperSoft e Talend Open Studio. Também será utilizado um mecanismo para de coleta de dados em dispositivos móveis. E por fim, serão apresentados relatórios e gráficos. A figura 9 ilustra o cenário proposto para o projeto da implementação do DW de Mandirituba (PR).

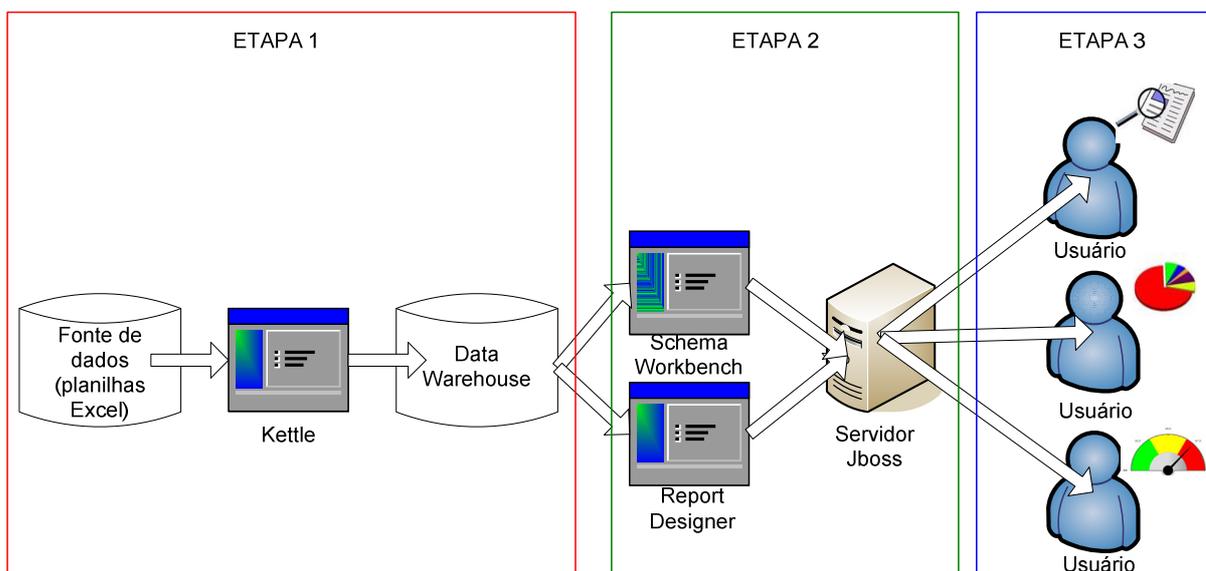


Figura 8 - Arquitetura proposta para o projeto Data Warehouse de Mandirituba (PR)
Fonte: Autoria própria.

A etapa 1 da figura 8 contempla o processo de ETL. A fonte dos dados está em formato do Excel. A etapa 2 contempla o processo de geração de relatórios, gráficos e *dashboard* e por fim a etapa 3 contempla a visualização pelos usuários das informações geradas pelas ferramentas da suíte Pentaho. Vale ressaltar, que apesar ter sido mencionado na figura 8 a ferramenta Kettle para realizar o processo de ETL serão também utilizadas as ferramentas JasperETL e Talend Open Studio conforme abordado no início deste capítulo.

Observando a figura 9 nota-se a evolução onde no primeiro momento existia dados uma planilha e, após a transformação, o resultado obtido um conjunto de banco de dados organizado e estruturado.

4.2 EXTRAÇÃO

A extração é o primeiro procedimento a ser realizado. Porém antes de chegar nesta fase é realizado uma avaliação conforme abordado no capítulo 2 seção 2.1 que detalha o processo de análise, levantamento e avaliação para construção de um projeto de DW. Após realizar esta avaliação a etapa de extração é a primeira a ser realizada, onde se tem os dados que foram coletados e os quais serão transformados antes de dar a carga no DW.

Para realizar este processo serão utilizadas três ferramentas: Kettle da suíte Pentaho, JasperETL e Talend Open Studio todas *open source* (software livre). Como o processo de extração é o mais trabalhoso, a idéia de utilizar três ferramentas é para verificar a diferença em questões de facilidade de utilização, desempenho entre outras questões.

4.2.1 Kettle

O capítulo 3 seção 3.1.1 abordou conceitos importantes sobre a ferramenta Kettle da suíte Pentaho usada para realizar o processo de ETL através das informações levantadas sobre o município de Mandirituba (PR). “Manipular dados de muitas maneiras fazendo matemática ou lógica, operações, aplicando funções de cadeia, agrupando por uma ou mais colunas, classificando dados” (PULVIRENTI; ROLDÁN, 2011). Para este propósito será utilizada a ferramenta Kettle usada para transformar estes dados que estão em planilhas para informações que serão armazenadas em um SGBD estruturado e organizado. Posteriormente serão realizadas as consultas, geração de relatórios, implementação de cubos de dados com base nessas informações, gráficos e *dashboard*.

O processo de ETL é complexo, pois trata-se de uma etapa onde os dados vão ser transformados em informações relevantes ao negócio. Esses dados na maioria das vezes estão em planilhas, sistemas legados entre outras fontes, onde se

torna necessário uma avaliação do que é realmente importante para armazenar no DW.

A ferramenta Kettle auxilia neste processo de ETL, basicamente a ferramenta necessita de uma entrada de dados e uma saída. Como os dados levantados sobre Mandirituba (PR) estão em formato de planilhas, este será o arquivo de entrada de dados já para saída será utilizado o SGBD MySQL. Como visto na figura 10 na coluna Origem dos dados nota-se que as informações das planilhas estão bem resumidas neste caso não existe a necessidade de realizar uma avaliação dos dados. Porém a necessidade de utilizar a ferramenta Kettle para realizar o processo de extração dessas informações para o SGBD conforme representado na figura 10 coluna Dados modelados que apresenta uma estrutura de banco de dados organizada em tabelas. A seguir na figura 10 demonstra o processo usando a ferramenta Kettle para realizar a extração dos dados contidos em planilhas para o SGBD.

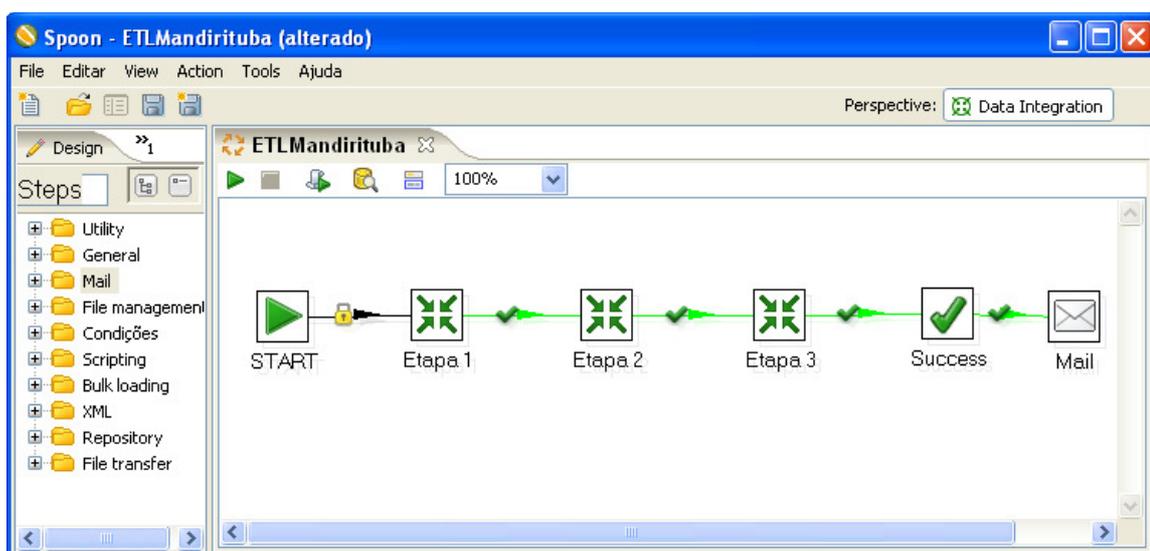


Figura 10 - Estrutura ETL realizado no Kettle
Fonte: Autoria própria.

A figura 10 apresentou o fluxo utilizado para dar carga às tabelas: tempo, consumo e classe que fazem parte da estrutura montada no SGBD MySQL para receber informações da planilha concedida pela Copel (Campania Paranaense de Energia Elétrica). Para iniciar o processo foi utilizado o Passo Start, após isso se inicia as etapas. A etapa 1 realiza o processo de carga na tabela tempo onde os dados são relativos a 10 anos de consumo de energia elétrica no município de Mandirituba (PR) iniciando no ano de 2000 até 2010. A etapa 2 realiza o processo

de carga na tabela consumo. A etapa 3 realiza o processo de carga na tabela classe do SGBD MySql. Por fim o processo se finaliza com o Passo Success e Mail usado para enviar um e-mail informado a situação da extração. Com isso as tabelas foram carregadas com informações contidas na planilha e estão prontas para serem utilizadas para geração de gráficos e relatórios. A figura 10 mostra apenas o processo de carga na tabela do SGBD MySql referentes a energia elétrica, porém as tabelas: educação, frota e empregos seguem o mesmo fluxo.

4.2.2 JasperETL

A ferramenta JasperETL auxilia no processo de ETL. Basicamente, a ferramenta necessita de uma entrada de dados e uma saída, como os dados levantados sobre Mandirituba (PR) estão em formato de planilhas, este será o arquivo de entrada de dados já para saída será utilizado o SGBD MySql. A figura 12 demonstra a processo utilizando a ferramenta JasperETL para realizar a extração dos dados contidos em planilhas para o SGBD.

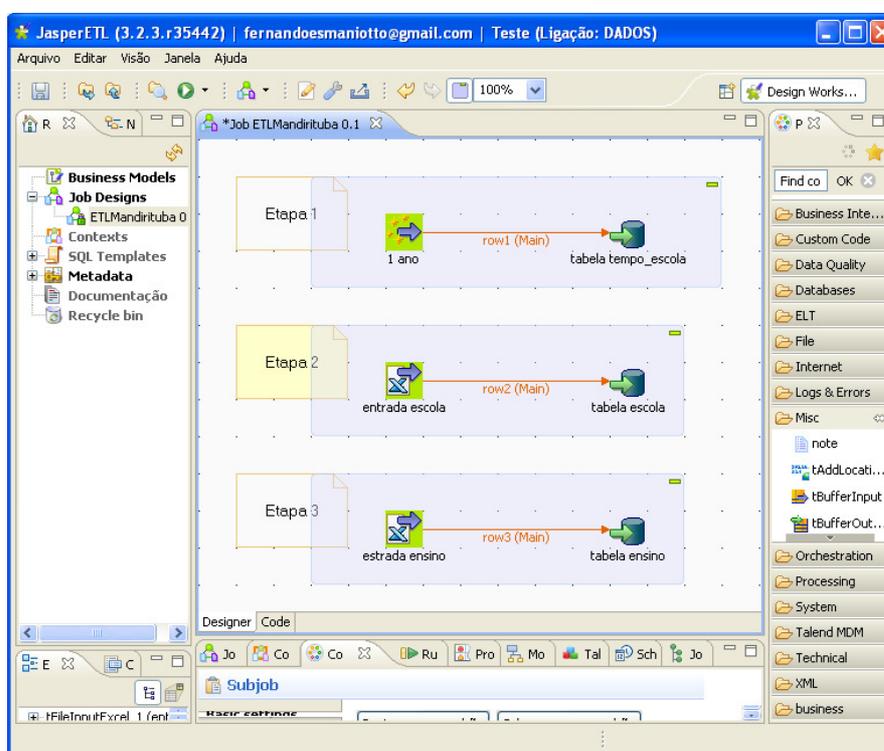


Figura 11 - Estrutura ETL realizada com o JasperETL
Fonte: Autoria própria.

A figura 11 mostra o fluxo utilizado para dar carga às tabelas: tempo, escola e ensino que fazem parte da estrutura montada no SGBD MySql para receber informações da planilha concedida pela Secretária de Educação da Prefeitura Municipal de Mandirituba (PR). A etapa 1 realiza o processo de carga onde os dados que vão povoar a tabela tempo do SGBD MySql são relativos a um ano da educação no município de Mandirituba (PR) referente ao ano de 2009. A etapa 2 realiza o processo de carga na tabela escola, onde os dados são referentes as escolas públicas, municipais e privadas da região. Por fim a etapa 3 realiza o processo de carga na tabela ensino, onde os dados de ensino são referentes a ensino fundamental, médio e pré escolar. Com isso as tabelas foram carregadas com informações contidas na planilha e estão prontas para serem utilizadas para geração de gráficos, relatórios entre outras atividades. Novamente existe uma facilidade para dar carga a tabela tempo, note na figura 11 na etapa 1 apresenta um fluxo relativamente breve se comparado com mesmo processo realizado com ferramenta Kettle.

4.2.3 Talend Open Studio

A partir deste ponto será apresentada a ferramenta Talend Open Studio usada para realizar o processo de ETL. Basicamente a ferramenta necessita de uma entrada de dados e uma saída, como os dados levantados sobre Mandirituba (PR) estão em formato de planilhas, este será o arquivo de entrada de dados já para saída será utilizado o SGBD MySql. A seguir na figura 12 demonstra a processo utilizando a ferramenta Talend Open Studio para realizar a extração dos dados contidos em planilhas para o SGBD.

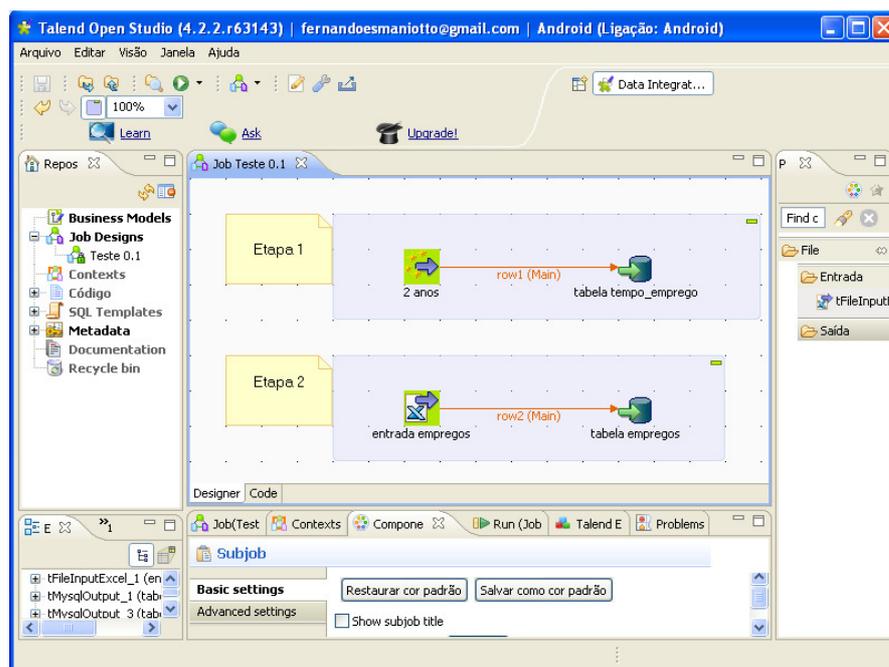


Figura 12 - Estrutura ETL realizada com Talend Open Studio
Fonte: Autoria própria.

A figura 12 mostra o fluxo utilizado para dar carga às tabelas: **tempo_emprego** e **emprego** que fazem parte da estrutura montada no SGBD MySql para receber informações da planilha empregos concedida pela Prefeitura Municipal de Mandirituba (PR). A etapa 1 conforme representada na figura 12, realiza o processo de carga onde os dados que vão povoar a tabela tempo do SGBD MySql são relativos a dois anos, referente a empregos no município de Mandirituba (PR). A etapa 2 realiza o processo de carga na tabela empregos, onde os dados são referentes a quantidade de empregos. Com isso as tabelas foram carregadas com informações contidas na planilha e estão prontas para serem utilizadas para geração de gráficos, relatórios entre outras atividades. Novamente existe uma facilidade para dar carga à tabela tempo, note na figura 12 etapa 1 onde o fluxo é relativamente breve referente ao mesmo processo realizado com ferramenta Kettle. Apesar de breve o resultado final é o mesmo.

4.3 COLETOR DE DADOS ANDROID

O estudo de caso Coletor de Dados Android apresenta a utilização da tecnologia Android para obtenção de dados. O sistema em questão permite coletar dados sociais para carregar a base dados com informações de campo. “Web

Services são um meio de expor uma tecnologia através da rede. Eles são um meio para chamar uma operação remota que não está amarrado a uma plataforma específica ou fornecedor e obter um resultado” (ABLESON; SEN; KING, 2011). O dispositivo faz conexão através de *WebService* para incluir, alterar, consultar e apagar dados em banco de dados.

A interface foi desenvolvida com base em dados levantados através da análise de requisitos feita na Secretaria Municipal de Ação Social de Mandirituba (PR), onde é utilizado um modelo de formulário impresso.

O sistema coletor de dados sociais oferece uma interface intuitiva onde praticamente as informações que o usuário precisa inserir são pré-definidas pelo sistema. A figura 13 ilustra a interface de cadastro de endereço.

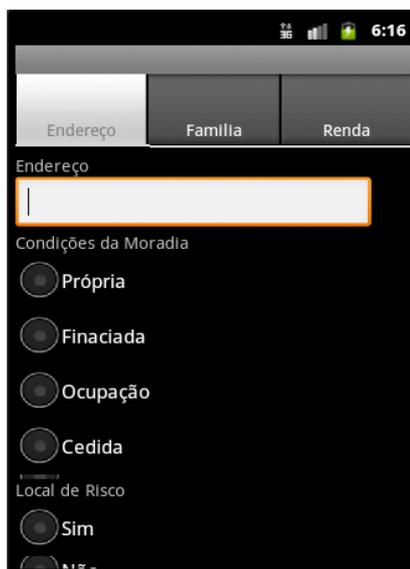


Figura 13 – Interface do coletor de dados sociais
Fonte: Autoria própria.

Nesta primeira interface existe apenas a apresentação de alguns itens que compõem o cadastro original de usuário da Secretaria Municipal de Ação Social de Mandirituba (PR). A idéia neste momento é mostrar a interface e o procedimento de envio das informações para uma base de dados e posteriormente o refinamento das informações que irão ser armazenadas no DW.

Para enviar as informações foi utilizada uma conexão com serviços para persistência de dados (*Web Services*). O servidor usado para armazenar o *Web Services* é Jboss e a tecnologia para transmissão de dados é o SOAP.

A figura 14 mostra o código de conexão com o serviço de persistência de dados, onde são enviados os dados para serem salvos no banco de dados através da conexão com *Web Services*:

```

public void enviar(){
    try {
        SoapObject envio = new SoapObject(NAMESPACE, OPERATION);
        SoapSerializationEnvelope envelope = new SoapSerializationEnvelope(SoapEnvelope.VER11);

        envio.addProperty("endereco", edtEndereco.getText().toString());
        if (rbAlugada.isChecked()) { envio.addProperty("condicaoMoradia", "Alugada"); }
        if (rbCedida.isChecked()) { envio.addProperty("condicaoMoradia", "Cedida"); }
        if (rbFinanciada.isChecked()) { envio.addProperty("condicaoMoradia", "Financiada"); }
        if (rbOcupacao.isChecked()) { envio.addProperty("condicaoMoradia", "Ocupação"); }
        if (rbPropria.isChecked()) { envio.addProperty("condicaoMoradia", "Própria"); }
        if (rbSim.isChecked()) { envio.addProperty("localRisco", "Sim"); }
        if (rbNao.isChecked()) { envio.addProperty("localRisco", "Não"); }

        envelope.setOutputSoapObject(envio);

        HttpTransportSE androidHttpTransport = new HttpTransportSE (SITE);
        androidHttpTransport.call(OPERATION, envelope);
        SoapObject resultado = (SoapObject) envelope.bodyIn;
        new AlertDialog.Builder (this)
            .setTitle("Envio de dados")
            .setMessage(resultado.getProperty(0).toString())
            .setNeutralButton("OK", new DialogInterface.OnClickListener() {
                public void onClick(DialogInterface dlg, int sumthin) { })
            .show();
    } catch (Exception e) {
        Log.e("log", e.getMessage());
    }
}
}

```

Figura 14 – Parte código de envio de dados
Fonte: Autoria própria.

Observando a figura 15 tem-se o código de conexão com o *Web Service*:

```

52  /**
53  * Operação de serviço web
54  */
55  @WebMethod(operationName = "salvarEndereco")
56  public String salvarEndereco (@WebParam(name = "endereco")
57  String endereco, @WebParam(name = "condicaoMoradia")
58  String condicaoMoradia, @WebParam(name = "localRisco")
59  String localRisco) {
60      facade = webFacade.getInstance();
61      facade.setEndereco (endereco);
62      facade.setCondicaoMoradia (condicaoMoradia);
63      facade.setLocalRisco (localRisco);
64      facade.salvar ();
65      return "Armazenado no servidor com sucesso!!!";
66  }
67
68
69
70 }
71

```

Figura 15 – Parte código do *Web Services*
Fonte: Autoria própria.

A imagem 15 mostra o código que salva os dados no servidor, ou seja, o cliente Android irá se conectar ao serviço chamado “**salvarEndereco**” repassando os respectivos dados. Através do aplicativo **MandiritubaMobile** torna-se possível uma integração de dados, onde se tem informações em tempo real para realizar o processamento.

4.4 RELATÓRIOS

Esta é a última parte do capítulo 4 a qual apresentada à ferramenta Pentaho Report Designer utilizada para gerar relatórios e gráficos. Através da utilização das ferramentas para realizar o processo de ETL, originou-se o DW de Mandirituba (PR). A partir deste ponto os dados que estão armazenados no DW serão utilizados para gerar relatórios e gráficos. Basicamente a ferramenta vai se conectar ao DW e a saída desta conexão serão em forma de relatórios.

A figura 16 mostra o resultado da conexão com a fonte de dados Educação contida no DW de Mandirituba (PR) em formato de relatório juntamente com gráfico outro recurso que a ferramenta oferece.

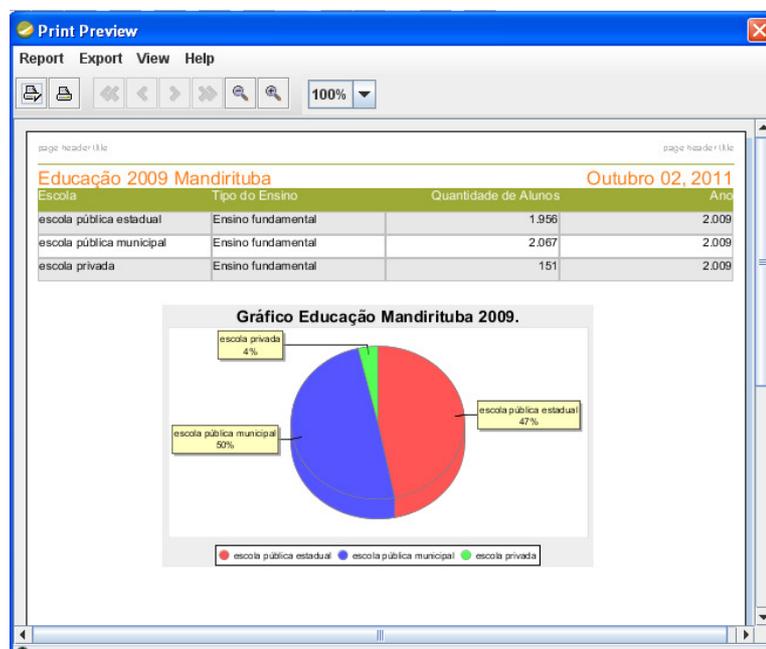


Figura 16 - Relatório e gráfico gerado pelo Pentaho Report Designer
Fonte: Autoria própria.

A figura 16 mostra um relatório com informações referentes ao ano de 2009 da educação municipal, estadual e privada do ensino fundamental. Apresentando a

quantidade de alunos matriculados neste período. Além do relatório é apresentado um gráfico que basicamente representa as mesmas informações contidas no relatório. A figura 17 apresenta apenas o gráfico, porém com algumas informações a mais do que o gráfico apresentado na figura 16.

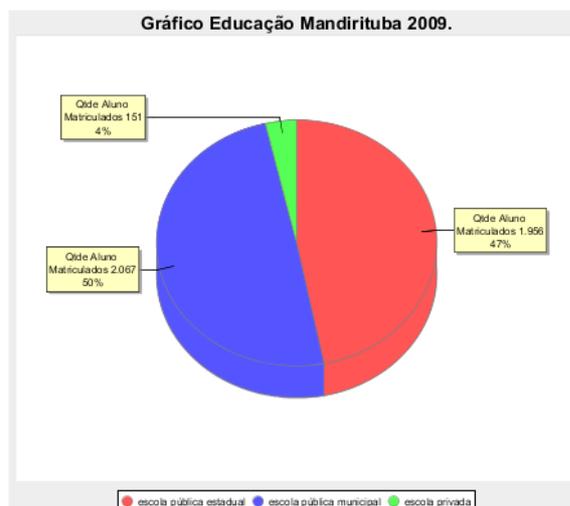


Figura 17 - Gráfico gerado pelo *Pentaho Report Designer*
Fonte: Autoria própria.

Basicamente foi incluída no gráfico representado na figura 17 a quantidade de alunos matriculados.

4.5 CONSIDERAÇÕES FINAIS

O capítulo 4 apresentou o processo de carga de dados do DW através de ferramentas ETL, coleta de dados com dispositivos móveis e também a apresentação de informações contidas no DW em forma de relatórios e gráficos. O próximo capítulo irá abordar mais resultados obtidos com a utilização das ferramentas da suíte Pentaho através de relatórios, gráficos e *dashboards* sobre o município de Mandirituba (PR).

CAPÍTULO 5 RESULTADOS

Os capítulos anteriores abordaram conceitos, regras e usabilidade das ferramentas além dos estudos de casos. Neste capítulo serão abordados os principais resultados obtidos através da suíte Pentaho. Através de relatórios, gráficos e *dashboard*, objetivando a importância de utilizar ferramentas para organizar informações e utilizar as mesmas para montar uma estratégia bem definida. A experiência com a ferramenta Pentaho Report Designer trouxe resultados em forma de relatórios e gráficos. O processo de utilização da ferramenta é prática e intuitiva oferecendo uma interface amigável para o usuário. Dentre suas funcionalidades destacam-se: geração de relatórios em vários formatos e geração de relatórios com gráficos.

5.1 RELATÓRIOS

Com a utilização da ferramenta Pentaho Report Designer foi gerado inúmeros relatórios. A figura 18 demonstra três relatórios sendo cada um deles possui um formato diferente, o primeiro está em formato PDF, o segundo em formato Excel e o terceiro em formato HTML.

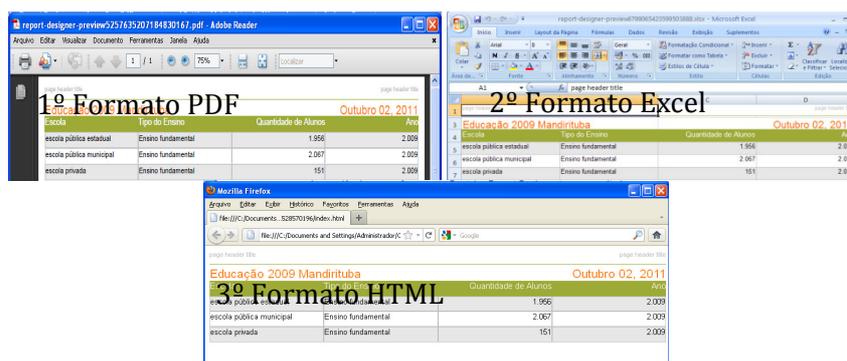


Figura 18 – Relatório da frota em formato: PDF, Excel e HTML
Fonte: Autoria própria.

A figura 18 apresenta o mesmo relatório, porém em formatos diferentes, além dos formatos apresentado na figura 18 existem outros, tais como: RTF, TXT e CVS, o que demonstra a versatilidade da ferramenta Pentaho Report Designer na geração de relatórios. A seguir será apresentado outros relatórios gerados tais como: Relatório de Consumo de Energia Elétrica, Relatório de Empregos e Relatório da

Frota de Veículos. A figura 19 mostra o relatório de consumo de energia elétrica do município de Mandirituba (PR), porém por haver um historio desde o ano de 2000 até o ano de 2010 armazenado no DW será apenas apresentado o consumo de energia elétrica do ano de 2010 da classe residencial.

Classificação	Mês	Ano	Consumo
residencial	janeiro	2.010	533
residencial	fevereiro	2.010	530
residencial	março	2.010	529
residencial	abril	2.010	562
residencial	maio	2.010	535
residencial	junho	2.010	525
residencial	julho	2.010	540
residencial	agosto	2.010	550
residencial	setembro	2.010	548
residencial	outubro	2.010	547
residencial	novembro	2.010	552
residencial	dezembro	2.010	547
0			

Figura 19 - Relatório do consumo de energia elétrica residencial
Fonte: Autoria própria.

Observando a figura 19 nota-se que no mês de abril houve um consumo maior com relação a os outros meses. A próxima figura apresenta um relatório sobre a frota de veículos do município referente ao ano de 2009.

Veículo	Quantidade	Ano
Automóvel	4.402	2.009
Caminhão	846	2.009
Caminhão trator	96	2.009
Caminhonete	620	2.009
Micro-ônibus	18	2.009
Motocicleta	1.381	2.009
Motoneta	195	2.009
Ônibus	52	2.009
Trator de rodas	4	2.009

Fim

Figura 20 - Relatório da frota de veículos
Fonte: Autoria própria.

Observando o relatório apresentado na figura 20 fica claro que a frota de automóveis é bem superior aos demais tipos de veículos. Por fim, a figura 21 representa um relatório da situação de empregos no ano de 2010 do município, observe a figura.

Mês	Quantidade de empregos
janeiro	218
fevereiro	256
março	323
abril	305
maio	358
junho	324
julho	345
agosto	313
setembro	299
outubro	290
novembro	287
dezembro	155

Figura 21 – Relatório sobre empregos
Fonte: Autoria própria.

Observando a figura 21 nota-se um aumento do número empregos nos mês de março, abril, maio, junho, julho e agosto.

Foram apresentado os principais relatórios obtidos através das informações contidas no DW de Mandirituba (PR). A próxima seção deste capítulo aborda a construção de gráficos.

5.2 GRÁFICOS

A ferramenta Pentaho Report Designer também oferece suporte para a implementação de gráficos. A partir deste ponto serão demonstradas informações em forma de gráficos, painéis de desenvolvimento além da visualização de informações através de dispositivos móveis sobre o município Mandirituba (PR).

Para dar início à série de gráficos à figura 23 apresenta um gráfico da frota de veículos de Mandirituba (PR), observe a figura.

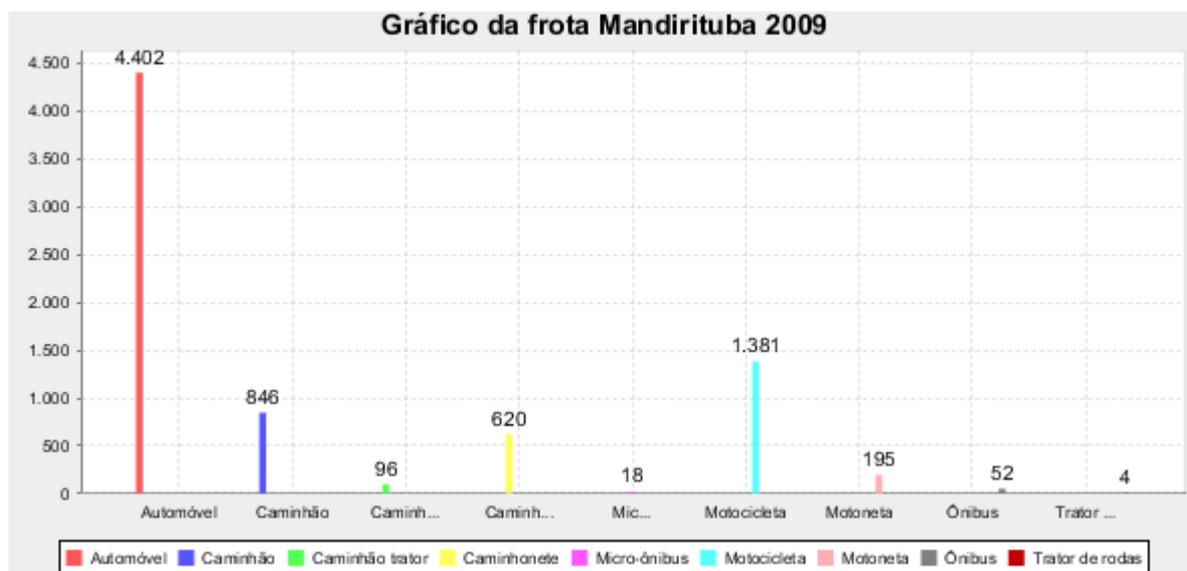


Figura 22 - Gráfico sobre a frota de veículos
Fonte: Autoria própria.

Observando o gráfico apresentado na figura 22 nota-se que o número de automóveis no ano de 2009 é bem superior aos demais tipos de veículos. O próximo gráfico apresenta os empregos demonstrados através de dois gráficos referentes aos anos de 2009 e 2010.

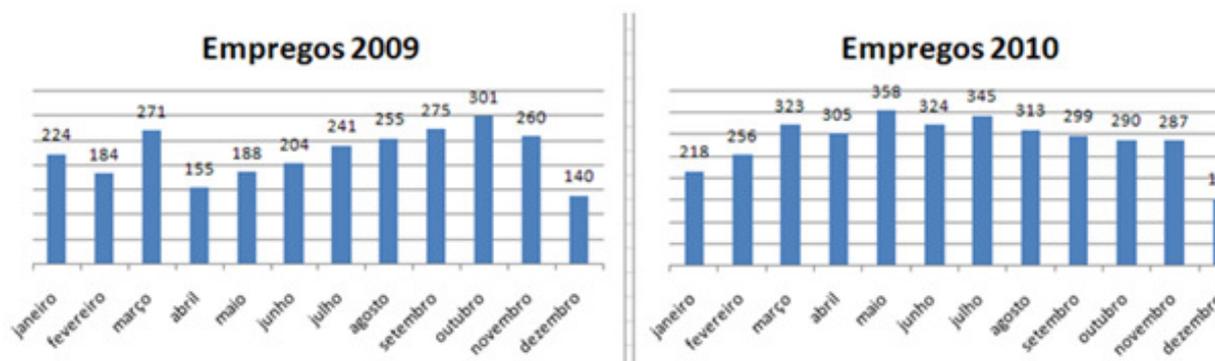
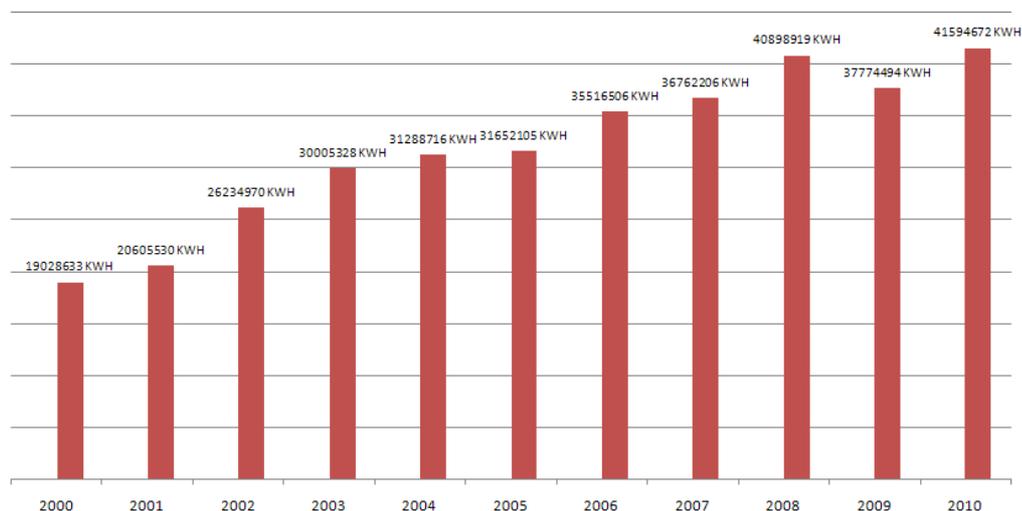


Figura 23 - Gráfico de empregos
Fonte: Autoria própria.

A figura do painel de empregos representado na figura 23 mostra um aumento de 12% no número de empregos em 2010. Isso quer dizer que a quantidade de pessoas empregadas que residem no município de Mandirituba (PR) obteve um aumento considerável se comparado com o ano de 2009. Também é possível observar que nos meses de março a agosto de 2010 obtiveram maior crescimento no número de pessoas empregadas.

O próximo gráfico apresenta o consumo de energia elétrica do município de Mandirituba (PR) dos últimos 10 anos nas classes: residencial, industrial, comercial, rural, poderes públicos, iluminação pública, serviços público e próprio. O ciclo se inicia no ano 2000 e se estende até o ano de 2010.



**Figura 24 - Gráfico do consumo de energia elétrica dos últimos 10 anos em Mandirituba (PR).
Fonte: Autoria própria.**

Através do gráfico apresentado na figura 24 é possível observar o aumento do consumo de energia elétrica na medida em que os anos vão passando, ou seja, cada ano o consumo de energia elétrica aumenta, note que no ano de 2010 o consumo se comparado com o ano de 2000 teve um aumento de 38%. Este aumento de leva em conta todas as classes de consumo que foram mencionadas anteriormente.

A ferramenta Pentaho Report Designer também gera relatórios e gráficos em formato HTML, com isso é possível integrar a um servidor de aplicação disponibilizando informações pela internet. O capítulo 4 seção 4.3 apresenta o estudo de caso do coletor de dados Android, cujo objetivo é de obter dados em tempo real sincronizando com um servidor e armazenando as informações em um banco de dados. Porém também é possível utilizar o recurso da mobilidade para obter informações em tempo real. A figura 25 apresenta o dispositivo Android obtendo um gráfico sobre informações referentes frota de veículos de Mandirituba (PR) referentes ao ano de 2009.

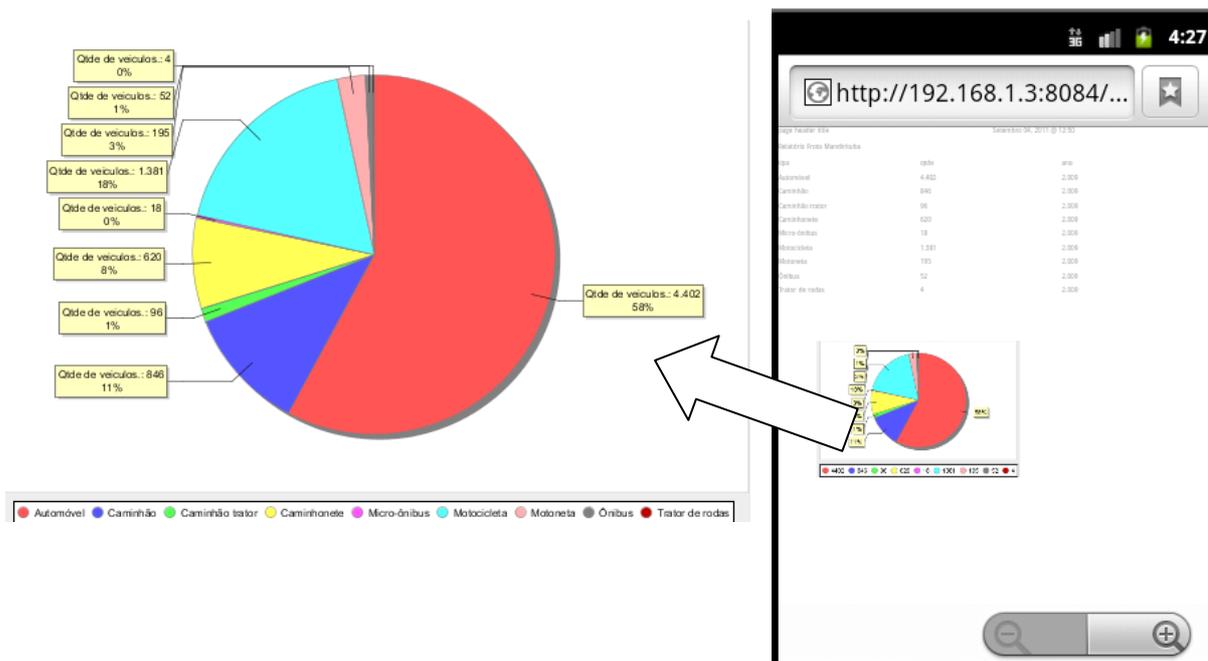


Figura 25 – Visualização de gráfico em dispositivo Android
Fonte: Autoria própria.

O recurso dos dispositivos móveis trouxe uma visão estratégica para o modelo do DW de Mandirituba-PR. No mesmo momento onde são coletas informações em campo através do coletor de dados sociais também é possível visualizar informações em tempo real.

5.3 DASHBOARD

“No contexto de Business Intelligence, um dashboard é uma aplicação usada para apresentar alto nível de conteúdo de BI para os usuários finais” (BOUMAN; DONGEN, 2009). Com os dashboards é possível visualizar informações em formato de painéis gráficos. O Framework CDF (Community *Dashboard* Framework) permite a implementação de *dashboard* dentro do Pentaho, a figura 26 apresenta dois *dashboard* do consumo de energia elétrica do ano de 2000 apresentando indicadores de consumo.

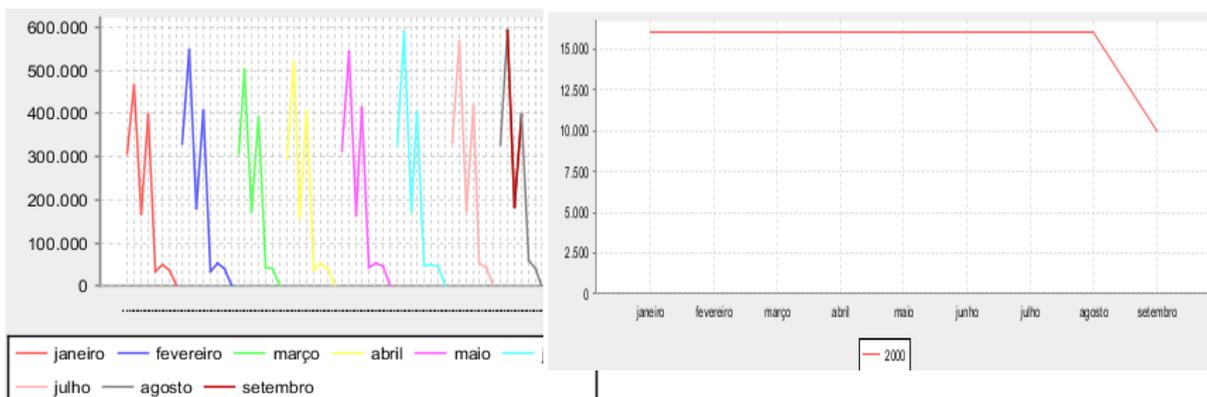


Figura 26 – Dashboard sobre o consumo de energia elétrica
Fonte: Autoria própria.

Visualizando a figura 26 que apresenta indicadores de consumo através de *dashboard* é possível verificar que no mês de setembro houve uma queda do consumo de energia elétrica no ano de 2000. O próximo dashboard apresenta indicadores de desempenho relacionados ao aumento do número de pessoas empregadas do município de Mandirituba (PR). Observe a figura 28.

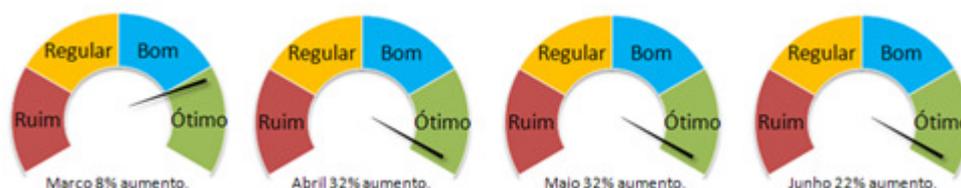


Figura 27 - Indicador do desempenho de empregos
Fonte: Autoria própria.

A figura 27 apresenta indicadores de desempenho que levam em conta os índices obtidos entre a comparação dos meses de março, abril, maio, junho de 2010 e 2009. Onde nos meses de março, abril, maio, junho foi obtido aumentos consideráveis com relação aos mesmos meses de 2009, destaca-se ainda os meses de abril e maio com aumento de 32% do número de pessoas empregadas.

5.4 CONSIDERAÇÕES FINAIS

Aplicar os recursos oferecidos pela suíte de aplicativos Pentaho operou de forma satisfatória sobre o projeto de *Data Warehouse* de Mandirituba (PR). Com base nas informações obtidas sobre o município foi gerado relatórios, gráficos e *dashboard*. Demonstrando aspectos relevantes de crescimento nos quesitos de emprego e consumo de energia elétrica do município.

Por outro lado, a utilização das ferramentas da suíte Pentaho pela visão de desenvolvimento, notou-se que o processo para implementação de um projeto de DW é contemplado totalmente pela suíte. Oferecendo suporte para processos críticos como o de processamento ETL, além dos recursos para geração de gráficos e relatórios. Resumindo a ferramenta é completa para realizar um projeto para implementação de DW.

CAPÍTULO 6 CONCLUSÃO E TRABALHOS FUTUROS

A ferramenta Pentaho propôs uma forma para construção de um ambiente de BI interativo e dinâmico. A suíte é composta de diversas ferramentas, as quais oferecem uma forma para integração, migração, obtenção de relatórios e gráficos, através de diferentes fontes de dados.

Esta monografia demonstrou algumas das funcionalidades do Pentaho. Ferramentas como Kettle para realizar extração e migração de dados. A ferramenta Mondrian servidor OLAP usado para visualização de cubos, Schema Workbench usado para implementação de cubos. Além das ferramentas da suíte Pentaho a monografia abordou também conceitos através de estudos de caso demonstrou a usabilidade das ferramentas Talend Open Studio e JasperETL utilizadas para processamento ETL. Mobilidade esta cada vez mais atuante entre a sociedade, a possibilidade de visualizar uma informação a qualquer momento faz com que dispositivos móveis tornem-se cada vez mais presentes no dia-a-dia. O projeto abordou conceitos de mobilidade para coleta de informações e visualização através de gráficos.

A monografia tornou-se fonte de grande aprendizado sobre uma tecnologia que aliada a estratégias bem definidas torna os negócios mais eficientes. A abordagem desta foi diretamente relacionada ao município de Mandirituba (PR) sobre dados sociais e informações da estrutura da cidade. Informações concedidas por órgãos como a Copel (Companhia Paranaense de Energia Elétrica) e a Prefeitura Municipal de Mandirituba permitiram que as informações demonstradas através de relatórios e gráficos apresentados no capítulo 5 sobre Resultados Obtidos, tornem-se uma fonte de conhecimento sobre o município.

A principal contribuição desta monografia foi no aspecto de centralizar informações, proveniente de órgãos estaduais e municipais em um único repositório de dados. E através dessa centralização de informações obter relatórios, gráficos, ou seja, informação de qualidade.

6.1 TRABALHOS FUTUROS

Certamente um ponto para executar um trabalho futuro diz respeito à questão da mobilidade. Este projeto apresentou a utilização de dispositivos móveis para coletar informações e também para apresentar em forma de gráficos. Porém existem diversos aspectos para melhorias. A seguir são apresentadas algumas sugestões:

- Desenvolvimento de sistema para visualizar gráficos usando Framework ZK e JPivot.
- Desenvolvimento de uma interface gráfica para visualizar gráficos e relatórios em dispositivos móveis.
- Desenvolvimento de uma interface gráfica para visualizar *dashboard* em dispositivos móveis.

REFERÊNCIAS BIBLIOGRÁFICAS

ABLESON, W. Frank; SEN, Robi; KING, Chris - **Android in Action Second Edition** – 2011.

BATINI, C., LENZERINI, M. **Comparative Analysis Of Methodologies For Database Schema Integration**, ACM Computing Surveys. New York, v.18, nº 4, pág.323 - 364, Dez/1986.

BOUMAN, Roland.; DONGEN, V. Jos - **Business Intelligence and Data Warehousing with Pentaho and MySQL** – 2009.

CHOPRA, Vivek; LI, Sing; GENENDER, Jeff – **Professional Apache Tomcat 6** – 2007.

GOLFARELLI, M. – **Open Source BI Plataforms: A Functional and Architectural Comparison. In: Data Warehousing and Knowledge Discovery**. Lecture Notes in Computer Science, 2009, Volume 5691/2009, 287-297, DOI: 10.1007/978-3-642-03730-6_23. Versão disponível em: <http://bias.csr.unibo.it/golfarelli/papers/DAWAK09%20-%20Golfarelli.pdf>. Acesso em: 16 de Setembro de 2011 às 15:00.

GONÇALVES, Marcio Elias. **Uma Ferramenta de Extração de Dados para Data Warehouse Baseada em Agentes Distribuídos**. Florianópolis: 2002. 64 p. Dissertação de Mestrado - Universidade Federal de Santa Catarina. Disponível em <http://www.datainfo.inf.br/marcio/download/artigos/mestrado.pdf>. Acesso em 10 de Setembro. 2011.

GORMAN, Will - **Pentaho Reporting 3.5 for Java Developers** – 2009.

INNON, Willian H - **Building the Data Warehouse fourth edition** – 2005.

JAMAR, Javid; JOHNSON, Peter – **Jboss In Action** – 2009.

JASPERSOFT - **JasperETL User Guide** – Disponível em: http://www.jaspersoft.com/sites/default/files/downloads/jbi4/4.1/etl/JaspersoftETL_Pro_UG_v41b_EN.pdf - Acesso em: 25 de Setembro de 2011 às 10:00.

PENTAHO ANALYSIS SERVICES: **Mondrian Project** – Disponível em <http://mondrian.pentaho.com/documentation.php> - Acesso em: 14 de Setembro de 2011 às 15:00.

PENTAHO REPORTING - **Introducing the Pentaho Report Designer** – Disponível em: <http://wiki.pentaho.com/display/COM/Introducing+the+Pentaho+Report+Designer> - Acesso em: 10 de Setembro de 2011 às 15:00.

PULVIRENTI, S. Adrián.; ROLDÁN, C. Maria - **Pentaho Data Integration 4 Cookbook** – 2011.

ROLDÁN, Maria Carina - **Pentaho 3.2 Data Integration Beginners Guide** – 2010.
TAKAOKO, Hiroo. **Marketing de Relacionamento no Varejo**. São Paulo-SP: Atlas, 1999.

TALEND – **Talend Open Studio User Guide** – Disponível em:
<http://www.talend.com/resources/documentation.php> - Acesso em: 25 de Setembro de 2011 às 09:00.

THOMSEN, E. **Construindo Sistemas de Informações Multidimensionais**. 2ª ed. São Paulo: Campus, 2002.

APÊNDICE A – Instalação da ferramenta Kettle

Para realizar o processo de instalação da ferramenta Kettle. Torna-se necessário efetuar download dos arquivos de execução do aplicativo. O download pode ser realizado na página:

<http://sourceforge.net/projects/Pentaho/files/Data%20Integration/4.1.0-stable/>. Feito o processo de download, basta descompactar o arquivo em algum diretório. “O único pré-requisito para instalar PDI é ter JRE 5.0 ou superior instalado” (PULVIRENTI, ROLDÁN, 2011). A figura 28 apresenta o diretório de instalação da ferramenta Kettle.

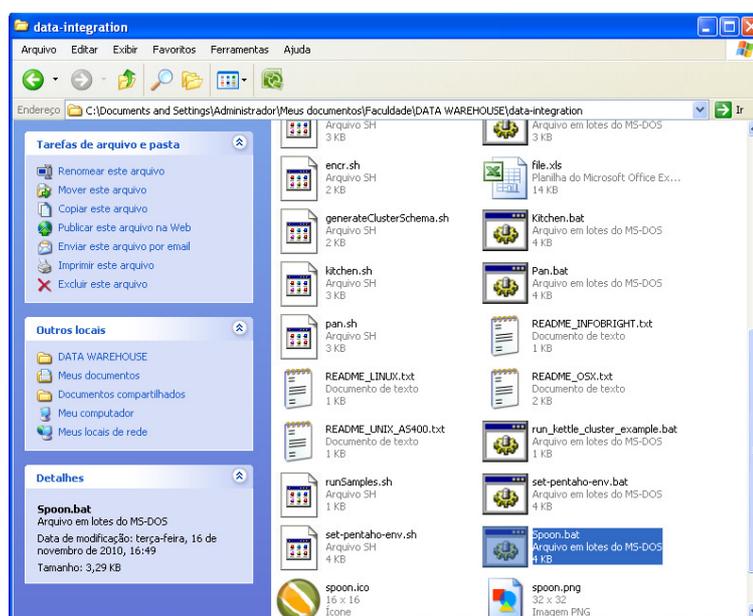


Figura 28 – Diretório de instalação do Kettle
Fonte: Autoria própria.

Ao clicar no ícone do Spoon será aberta a interface da ferramenta Kettle.

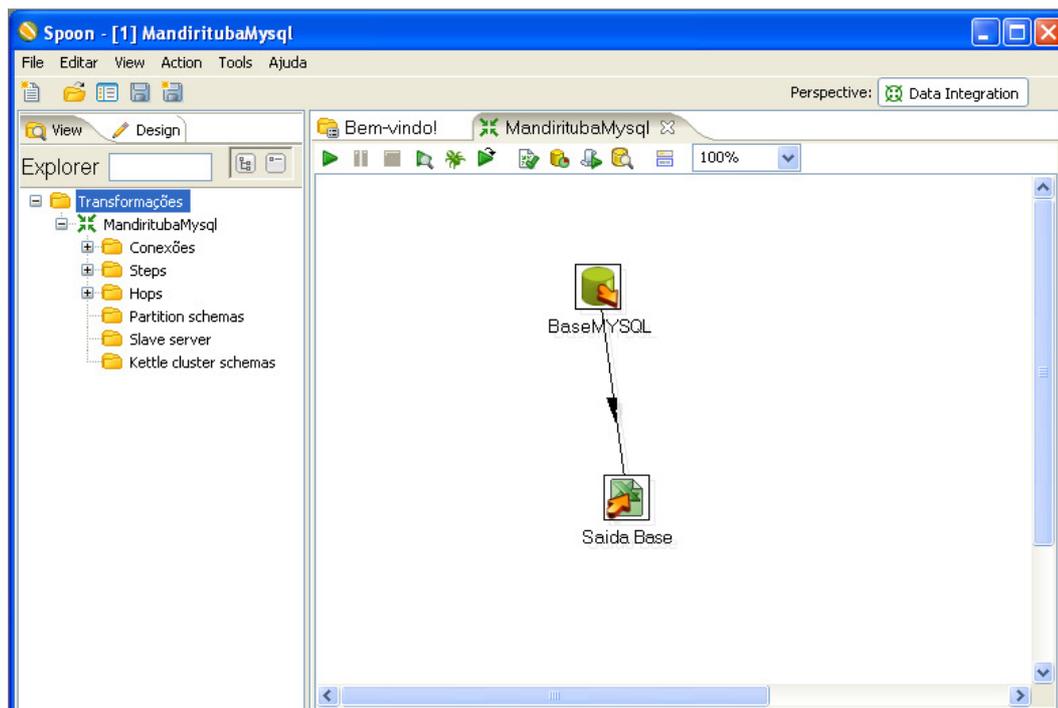


Figura 29 - Tela principal do Kettle
Fonte: Autoria própria.

APÊNDICE B – Instalação da ferramenta Talend Open Studio

Para realizar o processo de instalação da ferramenta Talend Open Studio é necessário realizar o download dos arquivos de instalação. Para isso, basta acessar a página: <http://www.talend.com/download.php>.

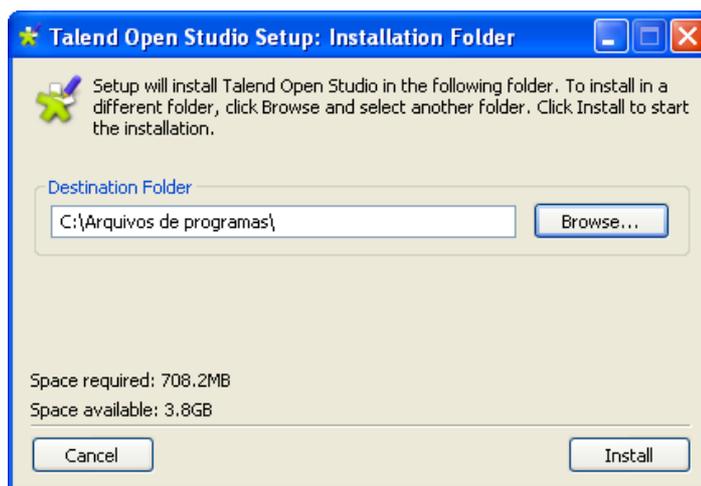


Figura 30 – Tela para escolha do diretório de instalação
Fonte: Autoria própria.

Ao clicar no botão Instalar será executado o instalador.

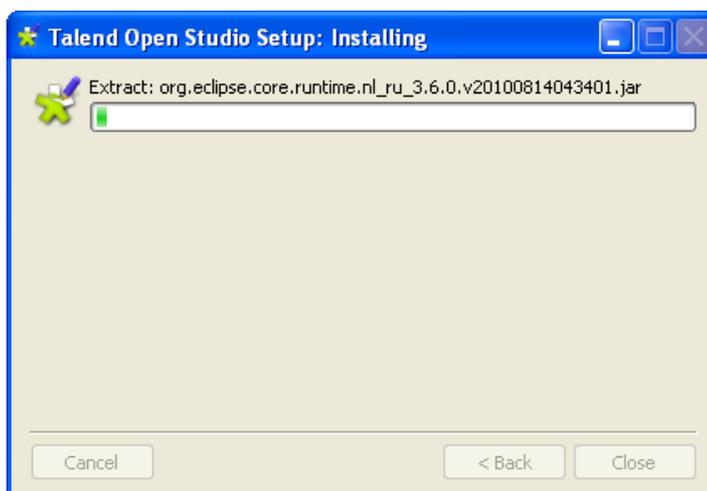


Figura 31 – Instalação sendo executada
Fonte: Autoria própria.

Ao executar o instalador é apresentada uma tela para confirmação de licença de uso. Para dar sequência à instalação é necessário clicar no botão **Aceitar**.

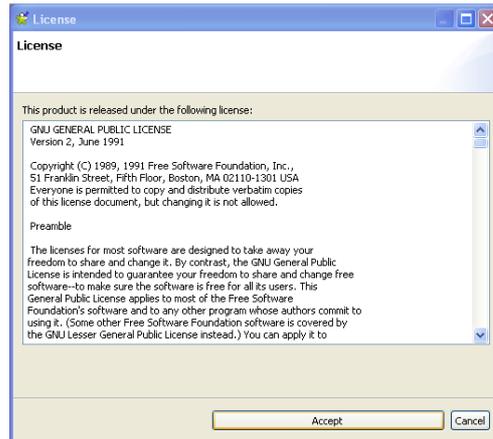


Figura 32 – Tela contrato de licença de uso
Fonte: Autoria própria.

A próxima tela é a de registro na comunidade. Para dar sequência a instalação é necessário clicar no botão **Registre mais tarde**.



Figura 33 – Cadastro na comunidade do Talend Open Source
Fonte: Autoria própria.

O próximo passo é a configuração do repositório local. Executando o aplicativo pela primeira vez deve-se clicar no botão  para realizar a configuração.



Figura 34 – Escolha do repositório
Fonte: Autoria própria.

Com isso tem a tela de configuração de repositório.

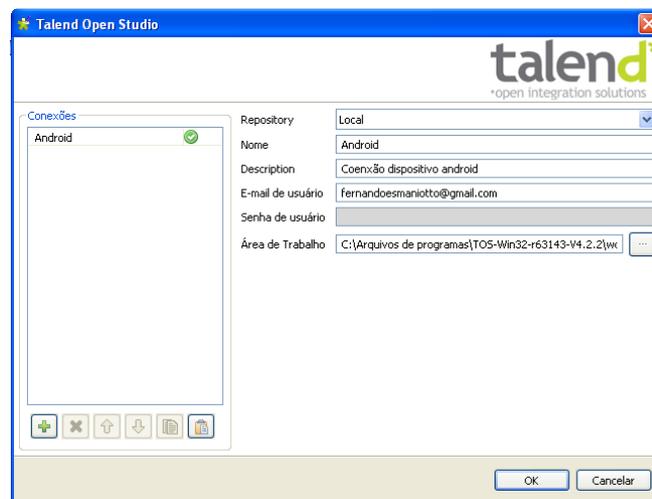


Figura 35 – Configuração do repositório
Fonte: Autoria própria.

Realizado a operação de configuração de conexão. É necessário escolher a opção **Criar um novo projeto local**. Com isso são carregadas as configurações que foram definidas para a conexão. Após isso basta clicar no botão **Vá!**.

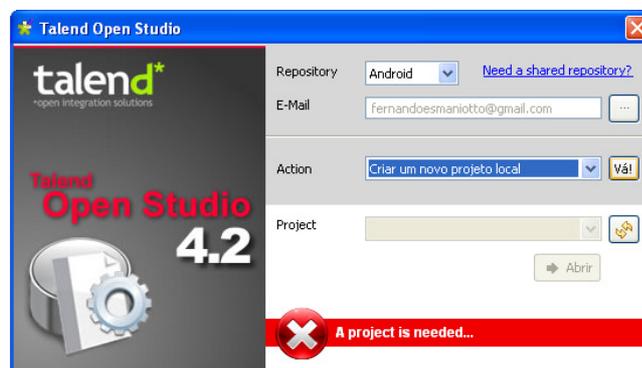


Figura 36 – Definição do projeto
Fonte: Autoria própria.

Na próxima tela é definido o nome para o projeto.

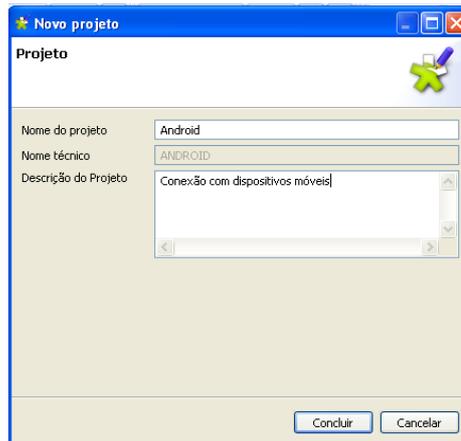


Figura 37 – Dados do projeto
Fonte: Autoria própria.

A figura 38 apresenta a tela inicial da ferramenta Talend Open Studio.

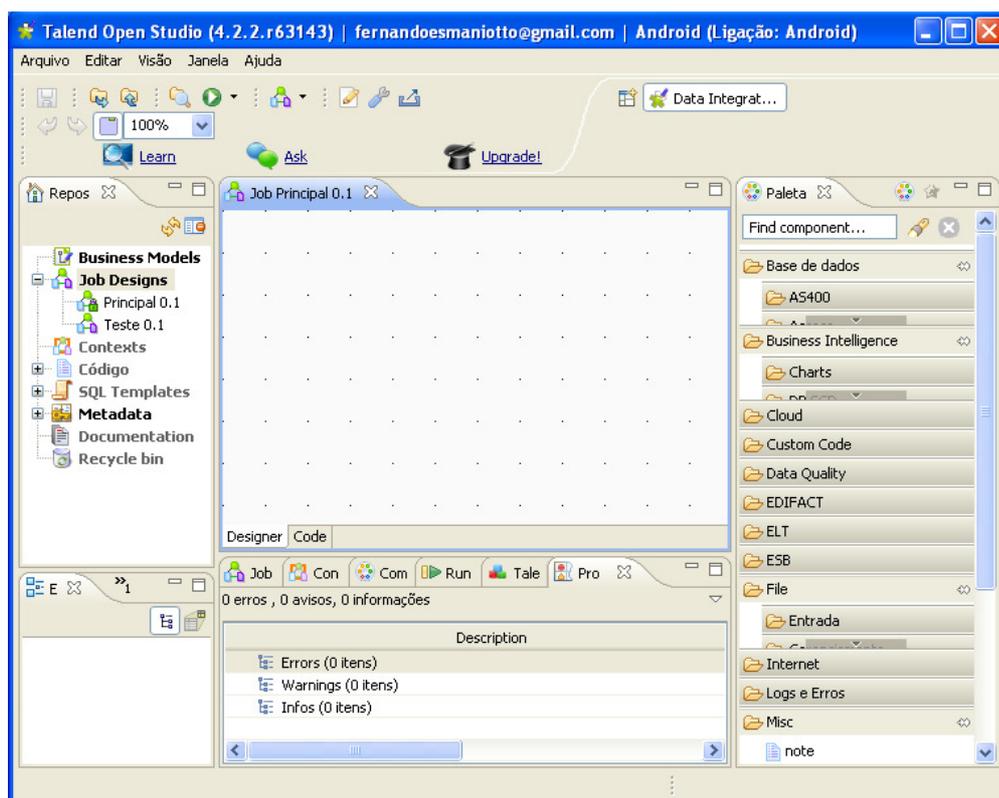


Figura 38 – Tela principal do Talend Open Studio
Fonte: Autoria própria.

APÊNDICE C – Instalação da ferramenta JasperETL

Para instalar a ferramenta JasperETL é necessário possuir o JVM 1.5 ou superior devidamente instalada no computador. Após realizar o download do arquivo na página: <http://sourceforge.net/projects/JasperETL>. Descompactar o arquivo em um algum diretório e iniciar o executável para a instalação.

Com a instalação concluída o próximo passo é realizar as configurações de conexão. Para isso clique no botão **Conexão** e escolha um e-mail. Finalizada a implementação da conexão será apresentado botões que permitem criar um novo projeto.



Figura 39 – Configuração de arquivos de demonstração
Fonte: Autoria própria.

APÊNDICE D – Instalação do servidor Mondrian

O primeiro passo a ser realizado para a instalação da ferramenta Mondrian é o download dos arquivos de execução. Para isso, basta acessar a página: <http://sourceforge.net/projects/Mondrian/>. Realizado o download, basta descompactar o arquivo em algum diretório.

Para que o servidor Mondrian funcione perfeitamente é necessário que o Apache Tomcat esteja devidamente instalado no computador. Para facilitar este processo efetue o download do Pentaho BI Server o qual vem vinculado com o Tomcat. O download pode ser realizado na página: <http://sourceforge.net/projects/Pentaho/files/Business%20Intelligence%20Server/>

Efetuada o download do Pentaho BI Server basta realizar o procedimento de descompactar os arquivos em algum diretório.

Para colocar em ação o Mondrian deve-se acessar a pasta **lib** que esta dentro da pasta Mondrian no diretório onde foi descompactado.

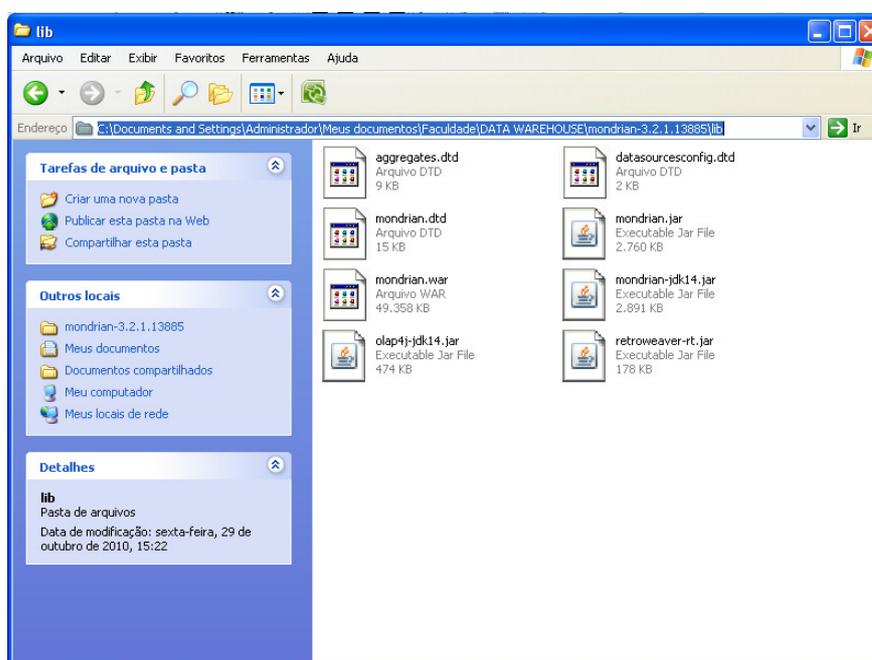


Figura 40 – Diretório onde se localiza o arquivo Mondrian.war
Fonte: Autoria própria.

Procure pelo arquivo **Mondrian. war**. É necessário copiar este arquivo para o diretório WebApps do Tomcat.

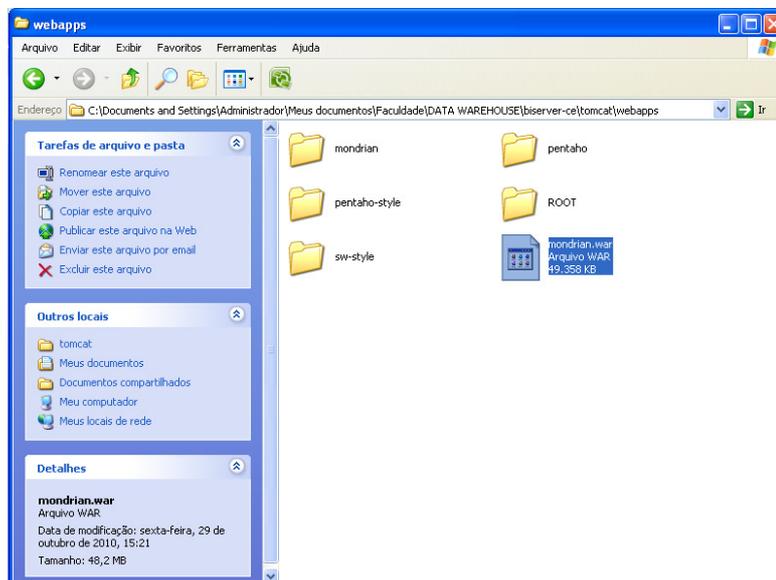


Figura 41 – Diretório webapps do Tomcat
Fonte: Autoria própria.

Após realizar o procedimento de configurações dos arquivos é necessário iniciar o Apache Tomcat. Para isso, basta acessar o diretório Bin dentro da pasta do Tomcat e localizar o ícone startup.bat e clicar duas vezes sobre ele. Realizado este procedimento o servidor Tomcat será inicializado.

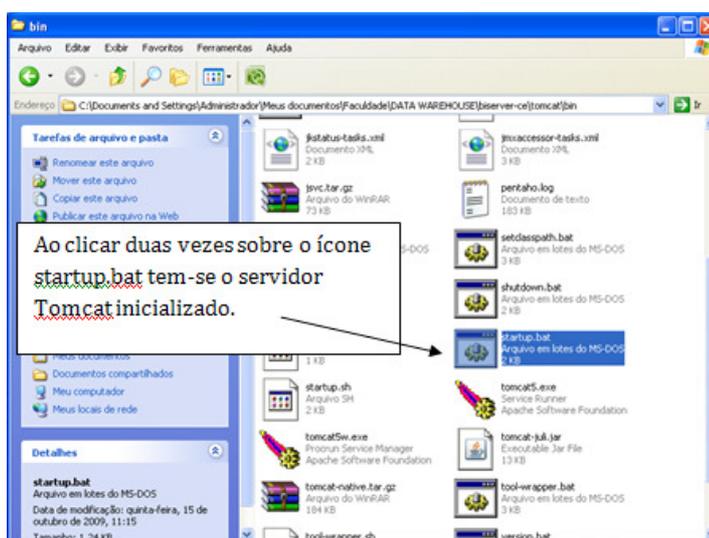


Figura 42 – Arquivo startup de inicialização do Tomcat
Fonte: Autoria própria.

Para ter acesso ao servidor Mondrian basta acessar a url: <http://localhost:8080/Mondrian/>. Logo será apresentada página principal do Mondrian.

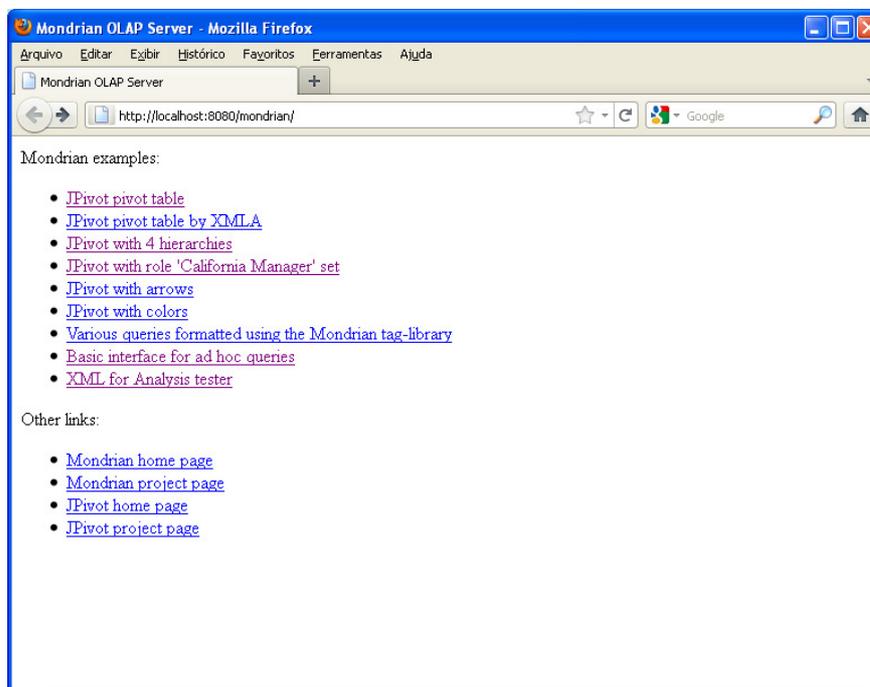


Figura 43 – Tela principal do servidor Mondrian
Fonte: Autoria própria.

APÊNDICE E – Instalação da ferramenta Schema Workbench

A ferramenta Schema Workbench é uma interface para implementação de Schemas. A seguir será apresentado o processo de instalação.

Após realizar o download dos arquivos de instalação do aplicativo na página: <http://sourceforge.net/projects/Mondrian/files/schema%20workbench/>, basta descompactá-lo em algum diretório.

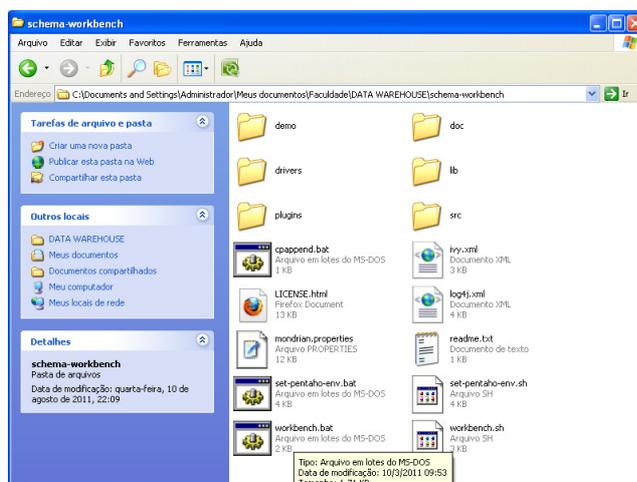


Figura 44 – Diretório de instalação do Schema Workbench
Fonte: Autoria própria.

Executar o arquivo workbench.bat para ter acesso a interface Schema Workbench.

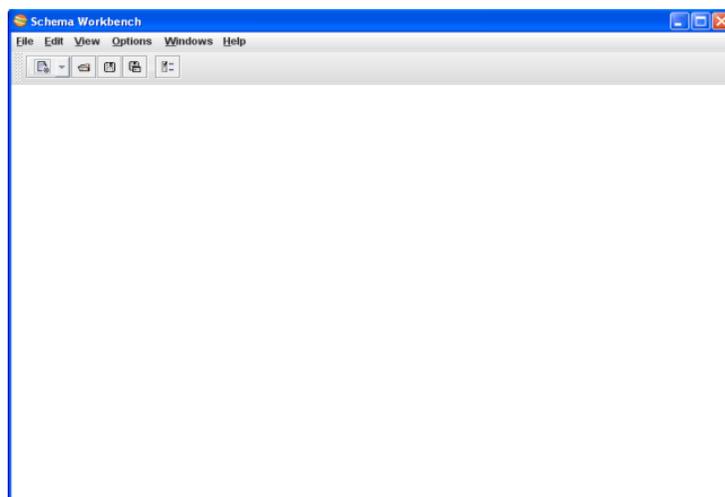


Figura 45 – Tela principal do Schema Workbench
Fonte: Autoria própria.

Para realizar o procedimento de criação de Schemas é necessário preparar uma conexão com banco de dados. Para realizar este procedimento é necessário acessar o menu **Opções/ Conexão**.

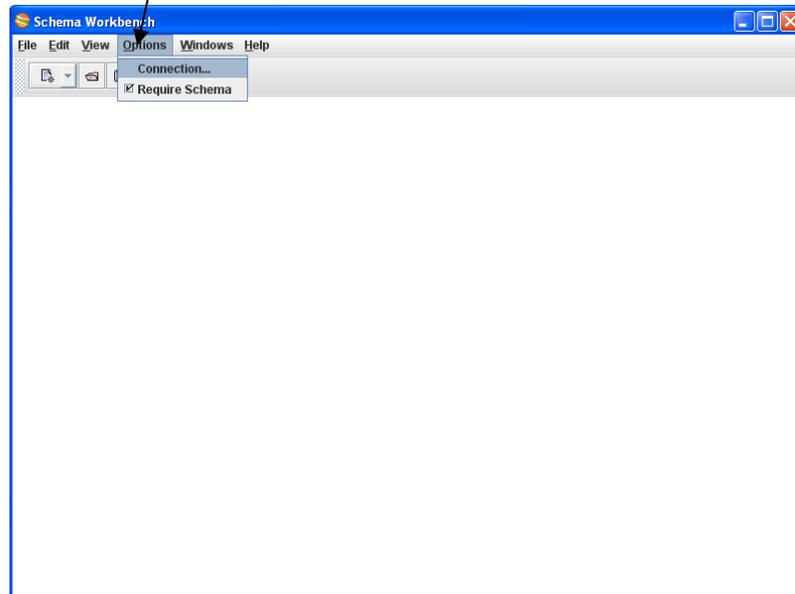


Figura 46 – Configuração da conexão com a base de dados
Fonte: Autoria própria.

Logo será apresentada a tela de **conexão com o banco de dados**.

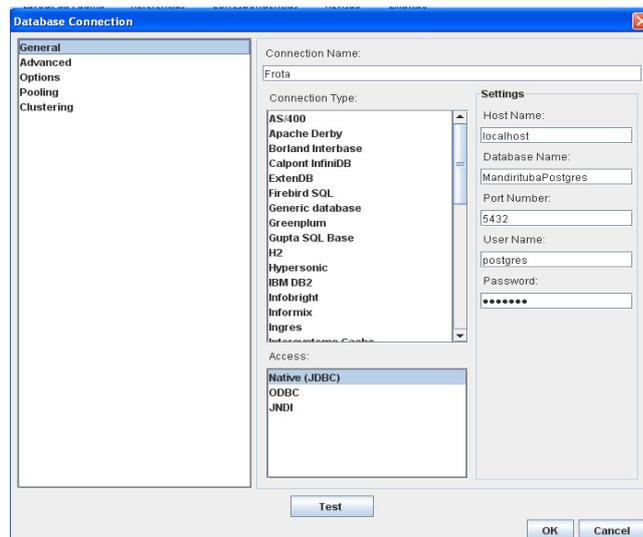


Figura 47 – Configuração da conexão com a base de dados
Fonte: Autoria própria.

Nesta etapa é necessário realizar a seguintes configurações:

- 1) **Tipo de Conexão:** Define o tipo do banco de dados;
- 2) **Nome da conexão:** Define o nome da conexão;
- 3) **Nome do Host:** Define o tipo de conexão local ou remoto;
- 4) **Nome da base:** Define o nome da fonte de dados;
- 5) **Número da porta:** Define a porta da conexão com a fonte de dados;
- 6) **Nome de usuário:** Define o nome de usuário para conexão;
- 7) **Senha:** Define a senha para acesso da fonte de dados.

Definindo os parâmetros, basta clicar no botão **Teste** para verificar se a conexão está funcionando corretamente.

APÊNDICE F – Instalação da ferramenta Pentaho Report Designer

Para instalar a ferramenta Pentaho Report Designer é necessário realizar o download dos arquivos de execução. Para isso, acessa a página: <http://reporting.pentaho.com/>. “Depois de baixar a distribuição binária do Report Designer, criar um diretório e descompactar o conteúdo do pacote para o diretório” (REPORT DESIGNER, 2011).

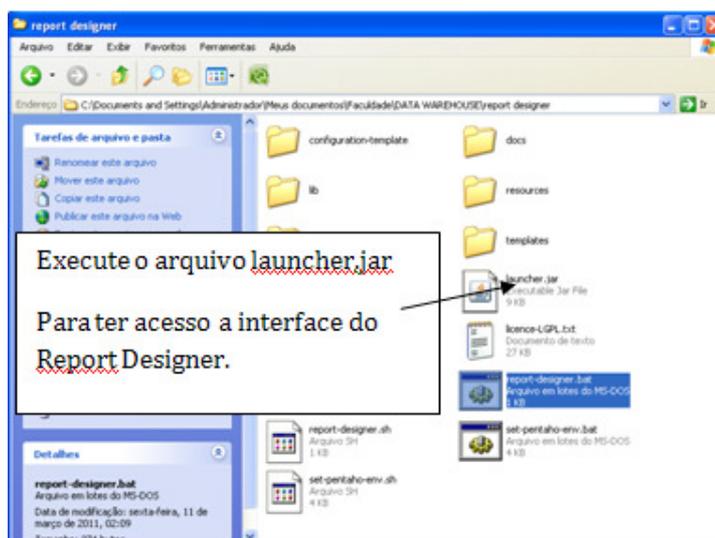


Figura 48 – Executando o arquivo Launcher
Fonte: Autoria própria.

Ao executar o arquivo tem acesso à interface do Report Designer.



Figura 49 – Tela principal do Pentaho Report Designer
Fonte: Autoria própria.

APÊNDICE G – Realizando a extração de dados com a ferramenta Kettle

“O Kettle oferece várias opções para obter informações de tabela de banco de dados” (PULVIRENTI; ROLDÁN, 2011). Para realizar o processo de extração de dados com a ferramenta Kettle será utilizada uma base de dados MySQL. Com o Kettle aberto, a primeira ação a ser tomada é a implementação de uma Transformação para tornar possível a execução das tarefas. Existem três formas para criar uma nova **Transformação**:

1. **Menu arquivo:** Opção **Novo / Transformação**;
2. Através do painel da esquerda que possui a opção Transformações.

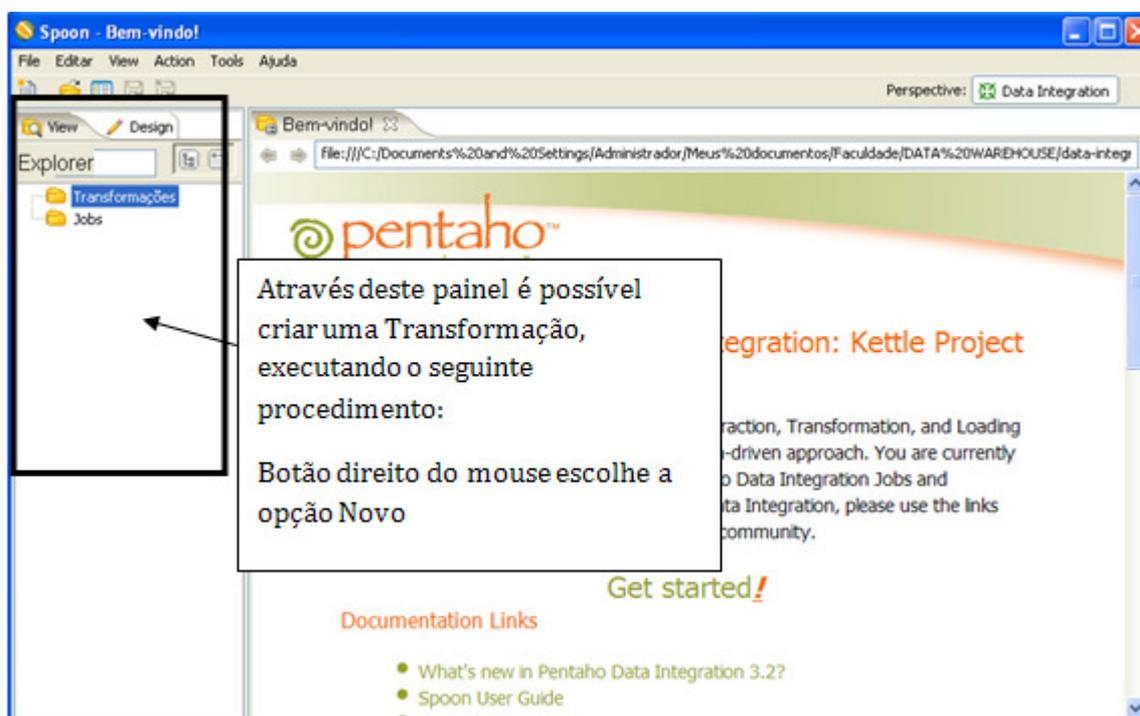


Figura 50 – Tela principal do Kettle
Fonte: Autoria própria.

2. **Usando teclado:** Efetuando a combinação das teclas CTRL-N.

Com a Transformação criada, basta salva-la.

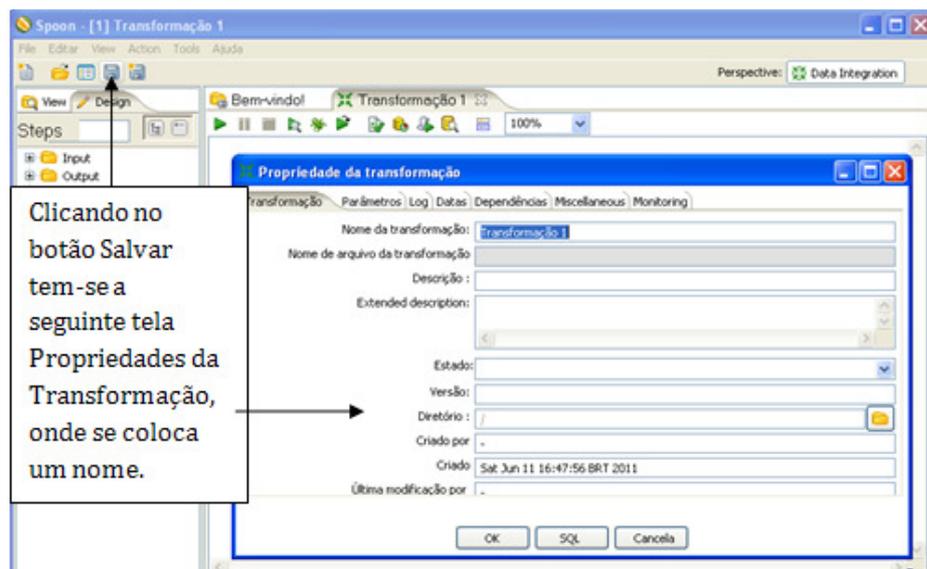


Figura 51 – Tela de propriedades da Transformação
 Fonte: Autoria própria.

Para inserir um novo componente primeiramente é necessário clicar no sinal de + do lado da pasta Input Input . Após isso escolher a opção Tabela de Input clique com o botão esquerdo do mouse e segure. Logo após, arraste até a área de trabalho do Kettle. A área em branco ao lado direito.

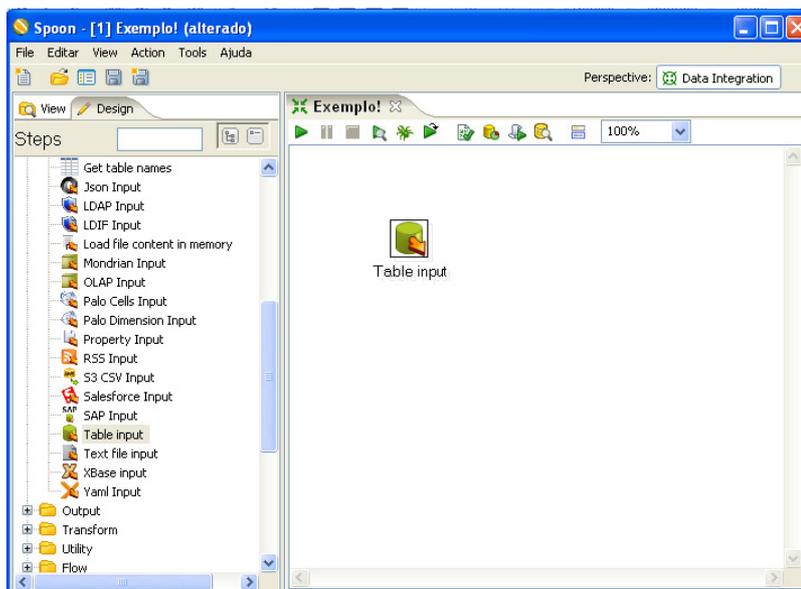


Figura 52 – Área de trabalho Kettle
 Fonte: Autoria própria.

O componente Tabela de entrada exige a configuração de conexão com banco de dados. Para isso clique sobre o **componente Tabela de entrada** duas vezes com o botão esquerdo do mouse para ter acesso à tela **Leitura de Tabela**.

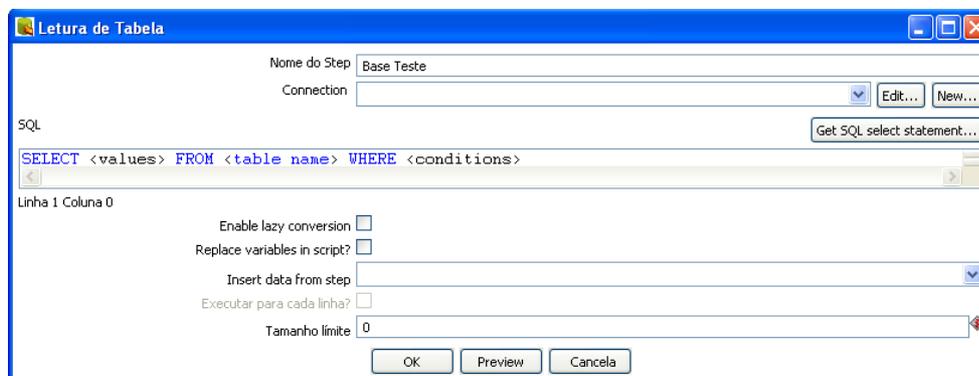


Figura 53 – Tela de Leitura de Tabela
Fonte: Autoria própria.

Nesta etapa são configurados a **conexão com o banco de dados** e as instruções SQL para consultas de dados. Para executar esta tarefa é necessário seguir três passos:

1. **Colocar o nome do Step:** Define um nome para o Step.
2. **Configurar uma conexão:** Configuração da conexão com banco de dados.

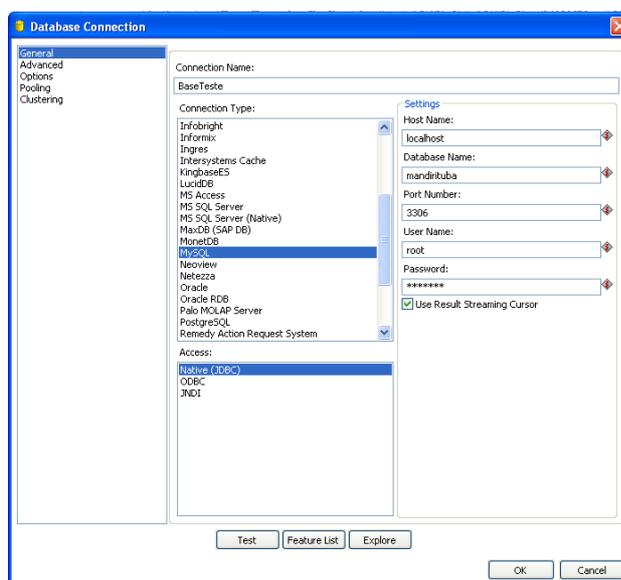


Figura 54 – Tela de conexão com a base
Fonte: Autoria própria.

Na tela de **Conexão com banco de dados** é necessário configurar sete itens, os quais estão relacionados abaixo:

1. **Nome da conexão:** Define o nome para a conexão
2. **Tipo da conexão:** Define o tipo do banco de dados que será utilizado na conexão.
3. **Acesso:** Tipo do acesso.
4. **Nome do host:** Define o local da base de dados.
5. **Nome da base:** Define qual é a base de dados utilizada.
6. **Nome de usuário:** Define o usuário do banco de dados.
7. **Senha:** Define a senha do banco de dados

Obs.: Tanto o **Usuário** quanto a **Senha** são definidos no momento da instalação do banco de dados no computador. Definidos os critérios, clicar no botão **Teste** para verificar a conexão.

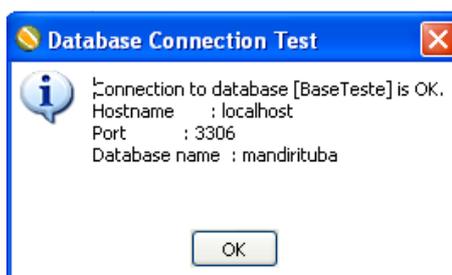


Figura 55 – Tela de Confirmação de conexão
Fonte: Autoria própria.

Comandos SQL: Ao colocar instruções SQL na tela de Leitura de Tabela na opção SQL. São definidos quais os dados a serem utilizados.

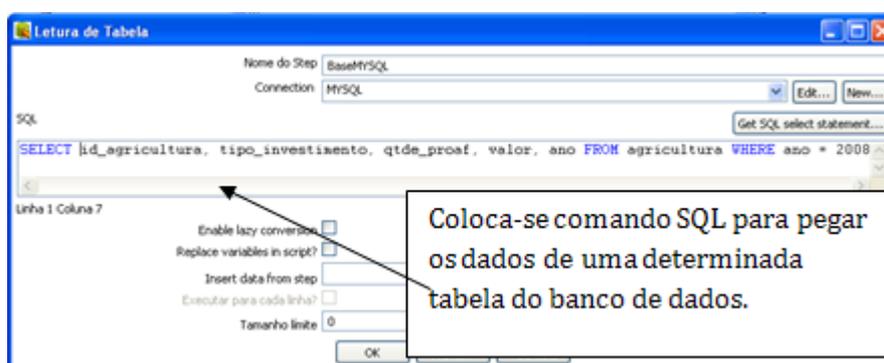


Figura 56 – Tela de leitura de tabela
Fonte: Autoria própria.

Após colocar os comandos SQL para **buscar os dados**, clica-se no botão OK. É necessário definir também a saída. Para isso, clica-se na opção Output. Logo

após isso pressiona o botão esquerdo do mouse sobre a opção **Excel Output** e arrasta para área de trabalho como feito anteriormente com a **Table Input**.

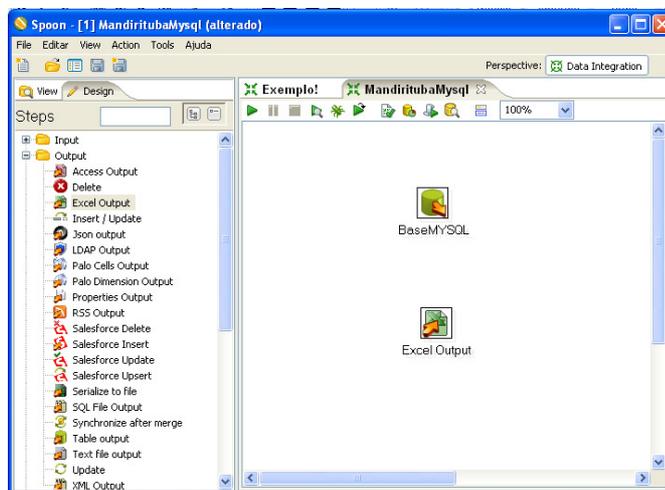


Figura 57 - Tela área de trabalho do Kettle
Fonte: Autoria própria.

No caso do componente de saída **Excel Output** não é realizado tantas configurações. Deve-se apenas, configurar o local onde o arquivo será salvo e também um nome para o componente. Para definir esses parâmetros, basta clicar duas vezes sobre o componente **Excel Output**.

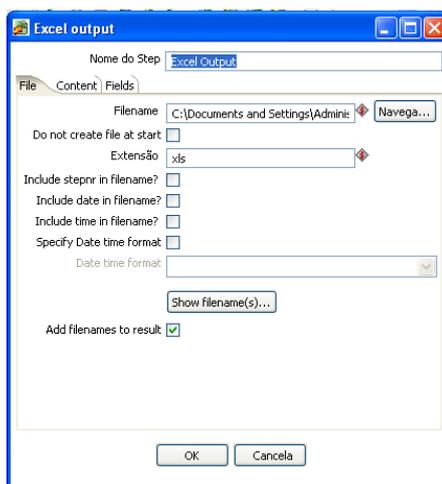


Figura 58 – Tela de configuração do arquivo de saída do Excel
Fonte: Autoria própria.

Em **Nome do Step** digita o nome para o Passo e em **Nome do Arquivo** o local onde será salvo. Definido os parâmetros basta clicar no botão OK. Ligar os

componentes de Entrada e Saída é o ultimo procedimento. Para isso, deve-se criar um **Hop**. Para criá-lo basta clicar com o botão Scroll sobre os componentes.

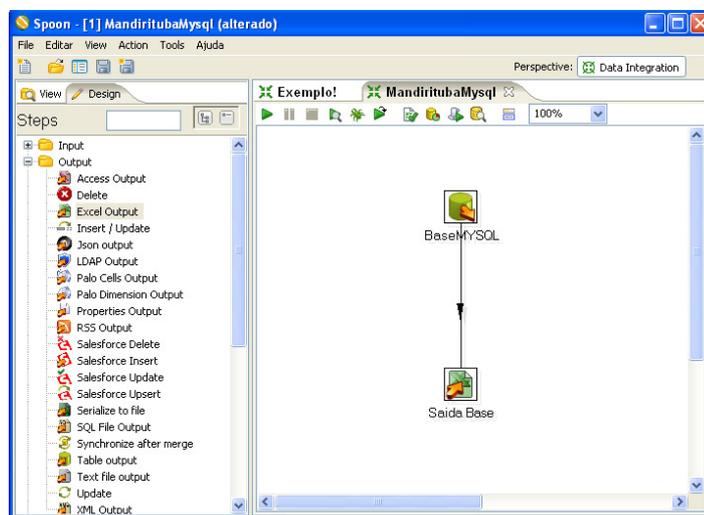


Figura 59 – Ligação entre a base de dados e arquivo Excel através do Hop
Fonte: Autoria própria.

Feito este procedimento, basta pressionar a tecla F9 para executar a **Transformação**. O resultado desta Transformação é a geração de uma planilha com dados que estão armazenados em banco de dados.

APÊNDICE H – Realizando a extração de dados com a ferramenta Talend Open Studio

Para realizar processo de ETL será utilizada uma base de dados MySQL de onde serão extraídas informações para um arquivo do Excel. Com o Talend Open Studio aberto, a primeira ação que deve ser tomada é a implementação de um Trabalho para executar as tarefas. Para isso, deve-se acessar o painel da esquerda que possui a opção Trabalho.

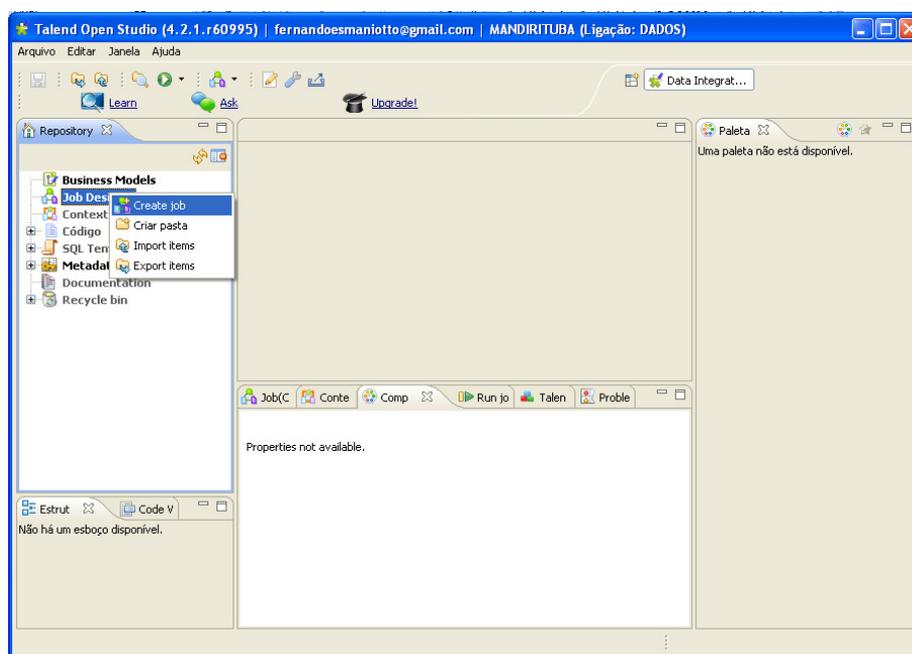


Figura 60 – Implementação do Job
Fonte: Autoria própria.

No momento que se clica com o botão direito do mouse sobre a opção Trabalho abre-se um menu **Pop-up** que possui a opção **Create Job**. Após escolher esta opção será apresentada a janela **New Job**.

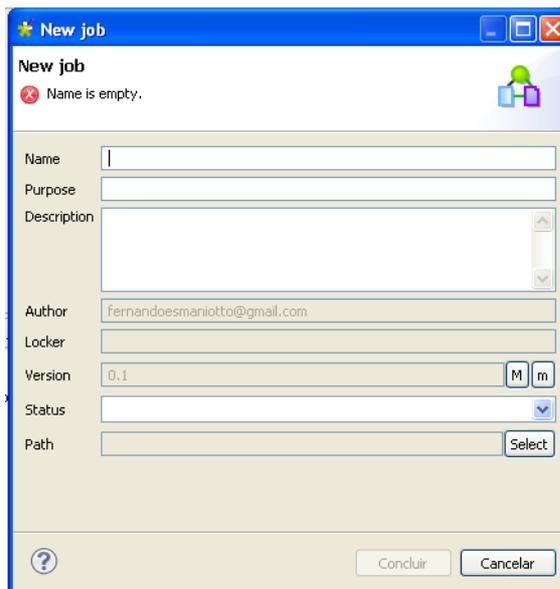


Figura 61 – Tela novo Job
Fonte: Autoria própria.

Após configurar o nome para o Trabalho, clica-se no botão Concluir.

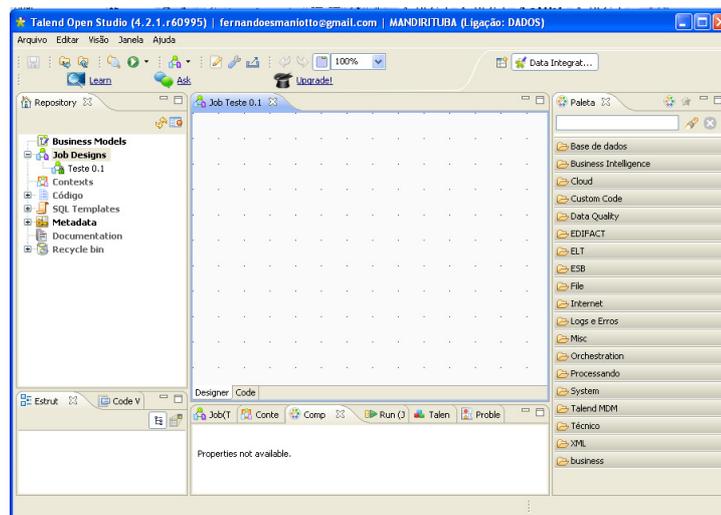


Figura 62 – Área de trabalho do Talend Open Studio
Fonte: Autoria própria.

Feito o procedimento de configuração do Trabalho. Basta incluir agora um objeto para **conexão com o banco de dados**. Para realizar este procedimento será utilizado o componente tMySQLInput_1, este componente é usado para realizar a conexão com banco de dados MySQL. Porém existe a necessidade de configurá-lo.

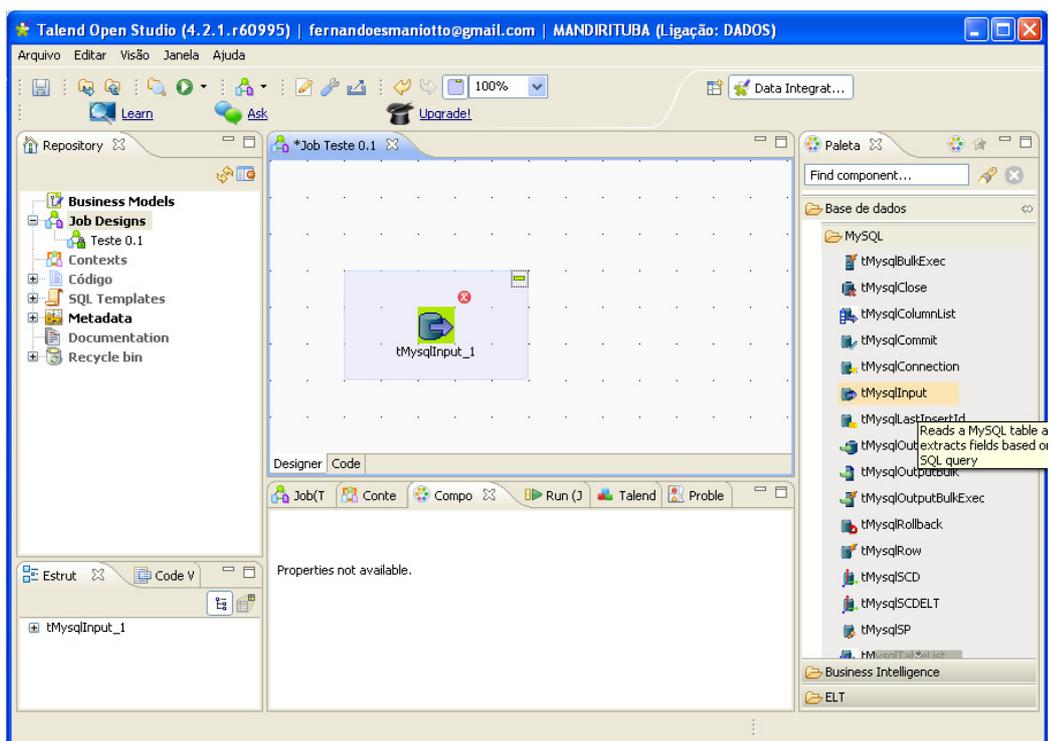


Figura 63 – Inserção do componente tMySqlInpu_1
Fonte: Autoria própria.

Neste momento é necessário realizar a conexão com o banco de dados. Para isso basta clicar sobre o elemento tMySqlInput_1 e na opção **Componente** localizada abaixo da área de design clicar sobre o ícone  para ter acesso ao **Assistente de conexão com banco de dados**.

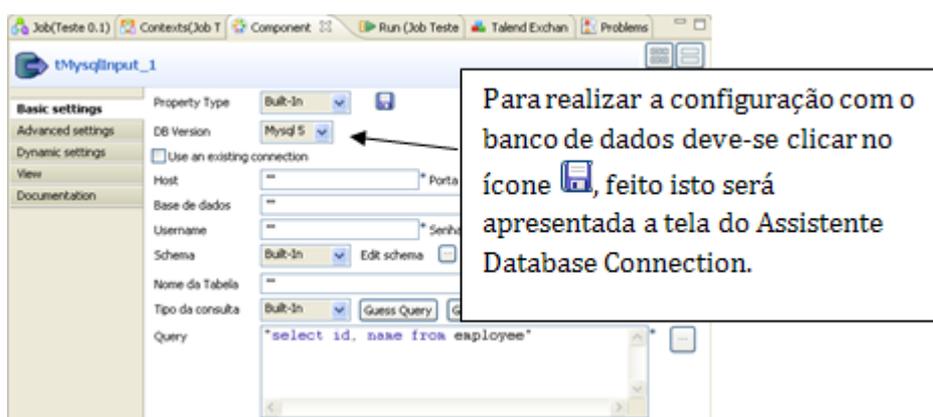


Figura 64 – Opção para realizar a conexão com a base de dados
Fonte: Autoria própria.

O Assistente de conexão com banco de dados é apresentado.

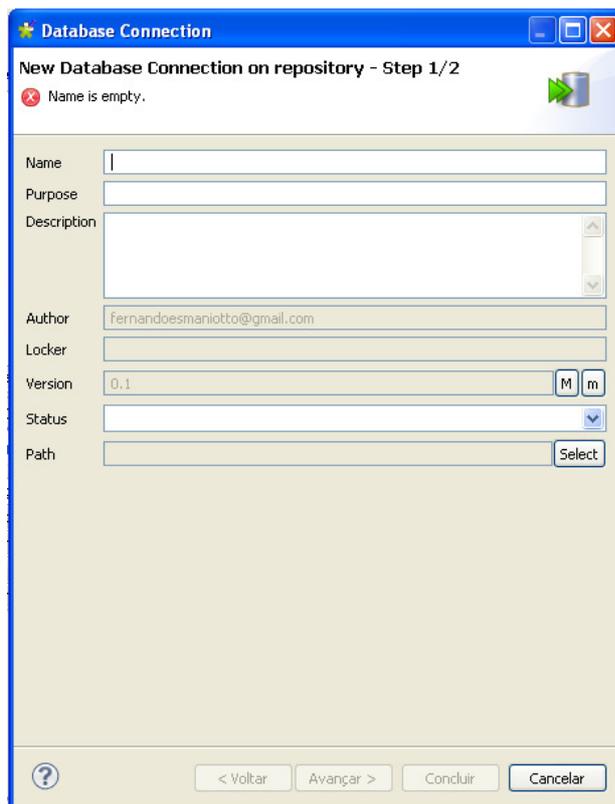


Figura 65 – Tela de configuração da base de dados
Fonte: Autoria própria.

O **Assistente de conexão com banco de dados** é dividido em duas etapas: na primeira é configurado um **nome para conexão**, também é possível colocar uma descrição para fins de documentação. Feito estas configurações clica-se no botão **Avançar**. Na segunda etapa do processo devem-se configurar os seguintes elementos:

1. **Tipo de banco:** seleciona qual o tipo de banco de dados para a nova conexão.
2. **Versão do banco:** seleciona qual a versão do banco de dados.
3. **Usuário:** define o nome de usuário do banco de dados.
4. **Senha:** define a senha do banco de dados;
5. **Servidor:** define o servidor que pode ser local ou remoto.
6. **Banco de dados:** define o nome da base de dados.

Obs.: Tanto o **Usuário** quanto a **Senha** são configurados no momento da instalação do banco de dados no computador.

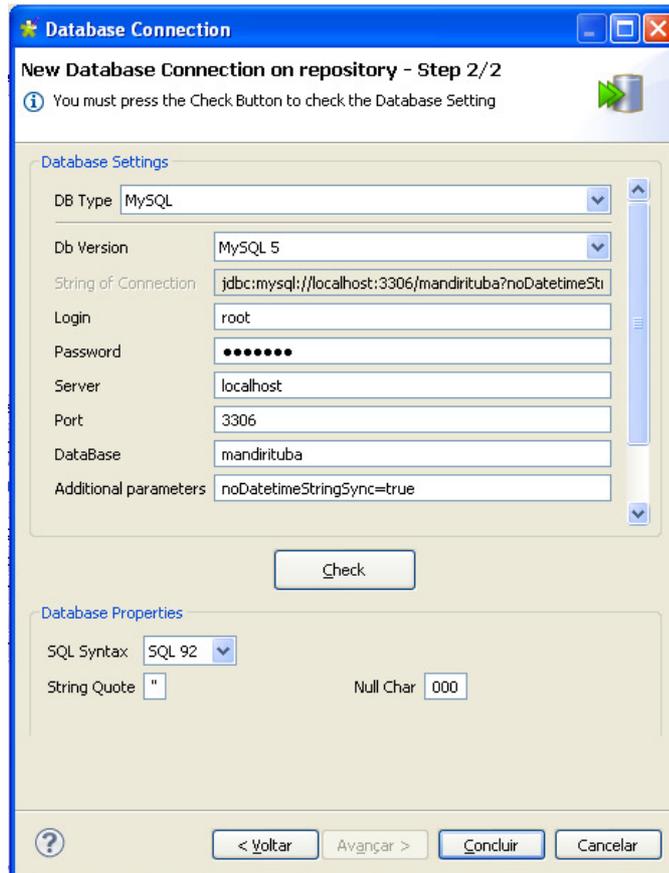


Figura 66 – Descrição das informações sobre a base de dados
Fonte: Autoria própria.

A figura 66 mostra os parâmetros que devem ser configurados na tela **Conexão com banco de dados**. Após realizar os procedimentos, basta clicar no botão **Concluir**. Antes disso pode-se clicar no botão **Verificar** para certificar que a conexão está funcionando.

Apesar da conexão com o banco de dados estar completa, ainda há a necessidade de configurar uma **Consulta para listagem dos dados**. Para realizar este procedimento deve-se acessar a aba **Componente** localizada logo abaixo da área de trabalho.

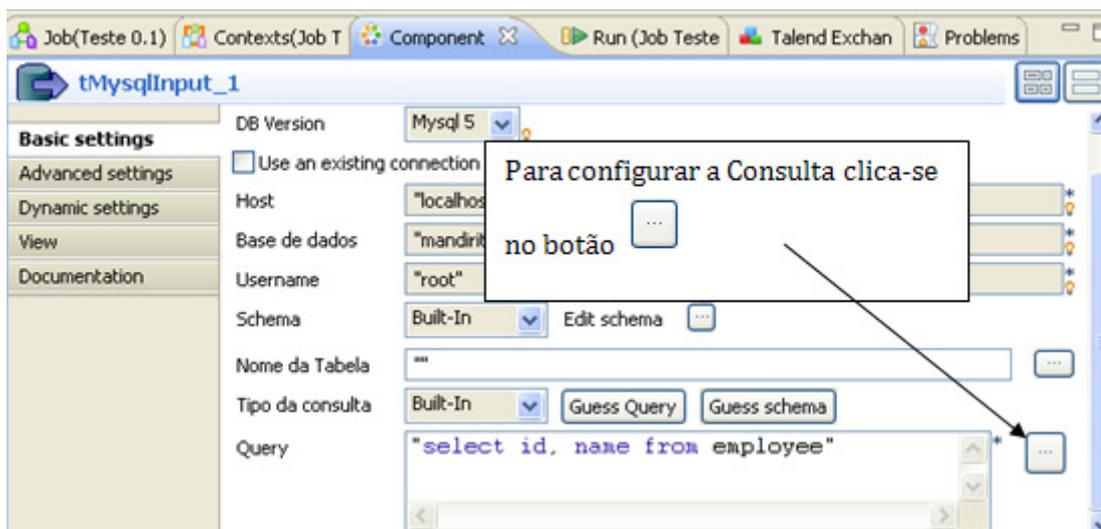


Figura 67 – Opção para realizar a configuração da consulta
Fonte: Autoria própria.

Clicando no botão  será apresentada a janela **SQL Builder**, nesta parte são executados três procedimentos.

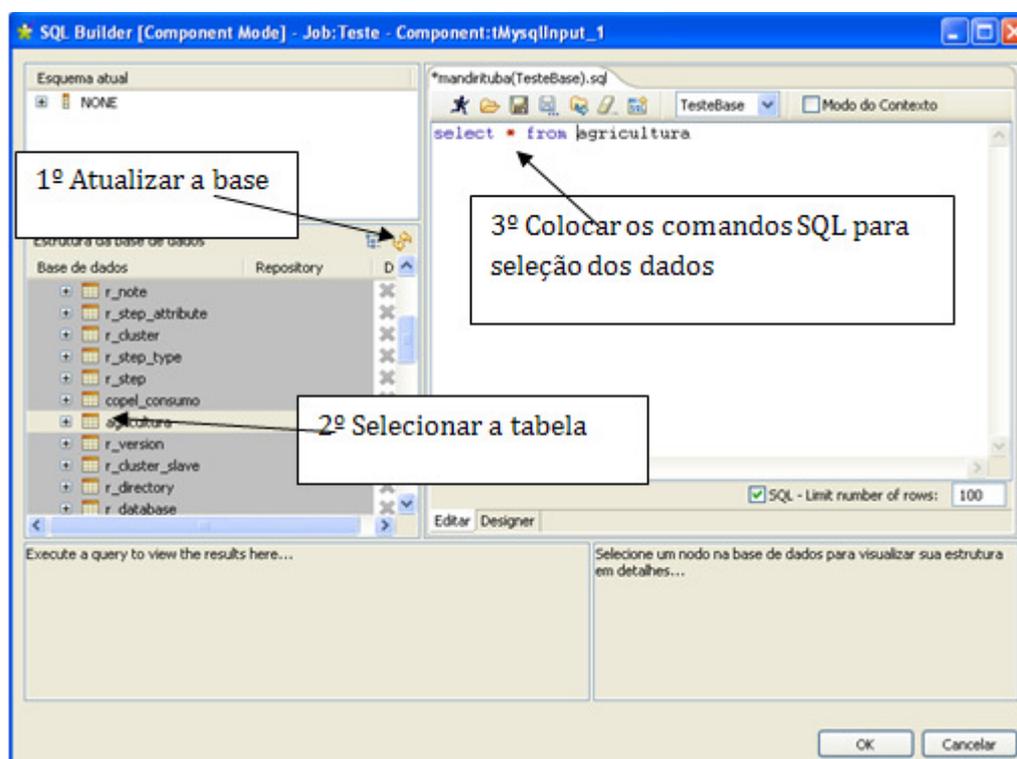


Figura 68 – Configuração da Consulta
Fonte: Autoria própria.

Logo então se clica no botão OK. Neste momento é adicionado o componente tFileOutputExcel_1 para realizar a saída. Este componente é encontrado na Paleta opção **File / tFileOutputExcel_1**.

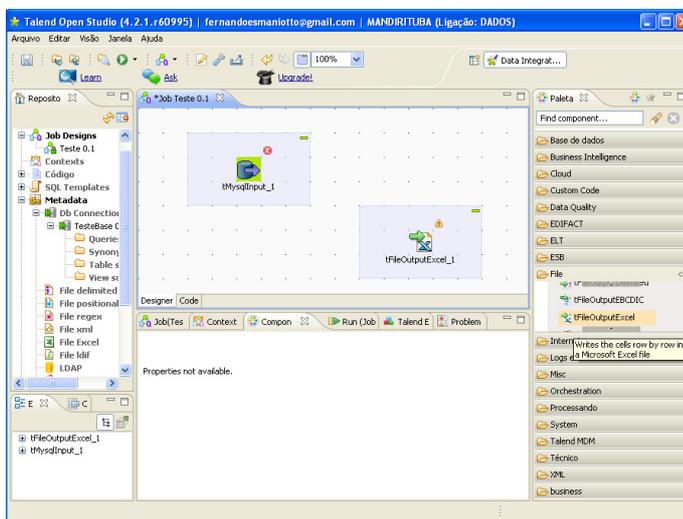


Figura 69 – Inserção do componente tFileOutputExcel_1
Fonte: Autoria própria.

No caso do componente TFileOutputExcel_1 não existe a necessidade de realizar tantas configurações, sendo apenas necessário colocar o local onde vai ser salvo o arquivo. Para isso, basta clicar sobre o componente e na aba **Componente** localizada abaixo da área de trabalho, escolher a opção **Nome do arquivo**, deve-se colocar o caminho do local onde será salvo o arquivo de saída.

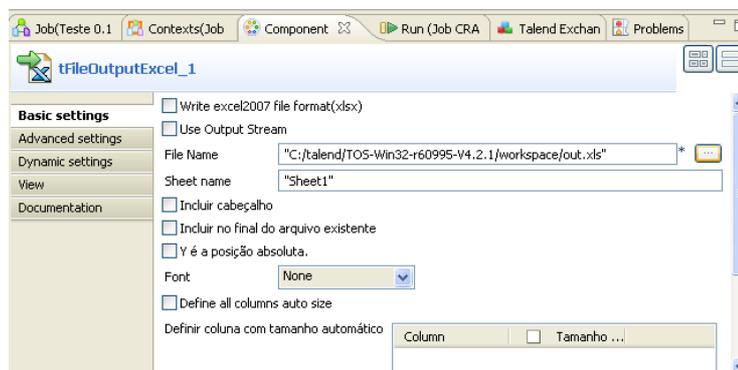


Figura 70 – Configuração do componente tFileOutputExcel_1
Fonte: Autoria própria.

Realizado os procedimentos têm-se os componentes de entrada e saída devidamente configurados. Porém falta ligá-los, para isso basta clicar com o botão

direito do mouse sobre o componente tFileMySQLInput_1 e arrastá-lo sobre o componente tFileExcelOutput_. Logo então, é possível executar o projeto.

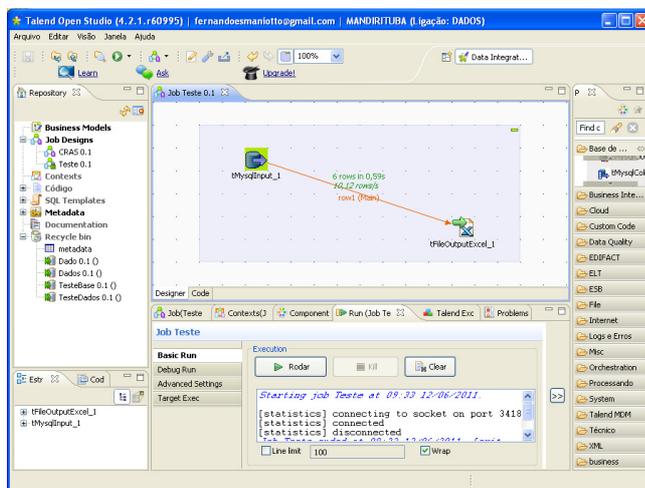


Figura 71 – Configuração da ligação entre base e arquivo de saída
Fonte: Autoria própria.

APÊNDICE I – Realizando a extração de dados com a ferramenta JasperETL

Para realizar o processo de ETL será utilizada uma base de dados MySQL de onde serão extraídas as informações, para um arquivo em formato Excel. Com o JasperETL aberto, a primeira ação que a ser tomada é a implementação de um Trabalho para executar as tarefas. Para realizar este procedimento deve-se acessar o painel da esquerda que possui a opção Trabalho.

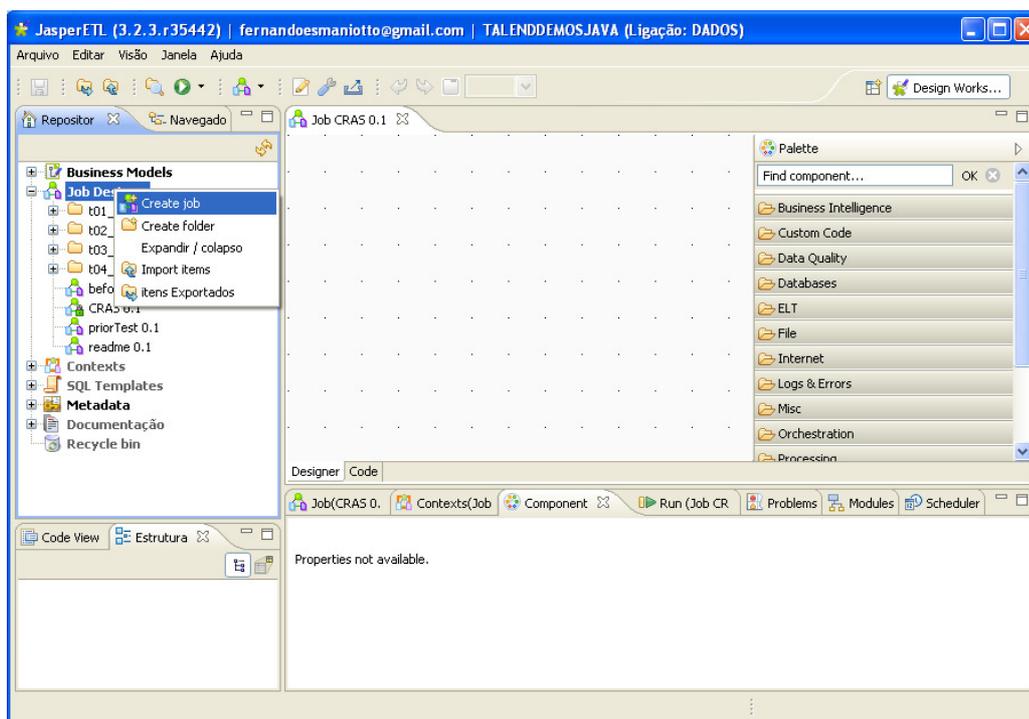


Figura 72 – Implementação do Job com JasperETL
Fonte: Autoria própria.

No momento que se clica com o botão direito do mouse sobre a opção Trabalho é aberto um menu **Pop-up**, o qual possui a opção **Create Job**. Após escolher esta opção será apresentada a janela **New Job**.

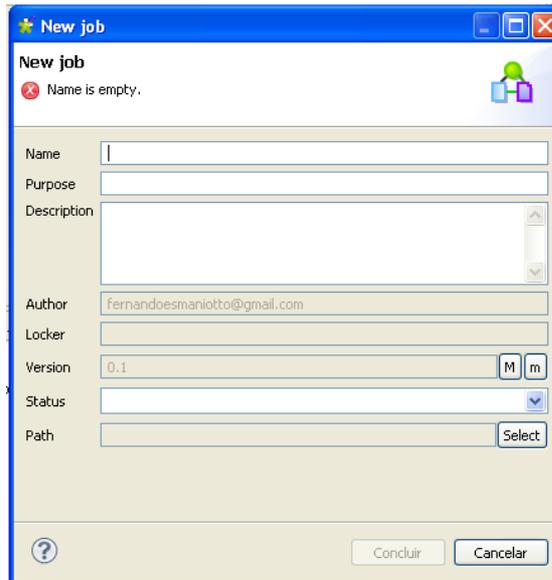


Figura 73 – Tela novo Job
Fonte: Autoria própria.

Após configurar o nome para o Trabalho clica-se no botão Concluir.

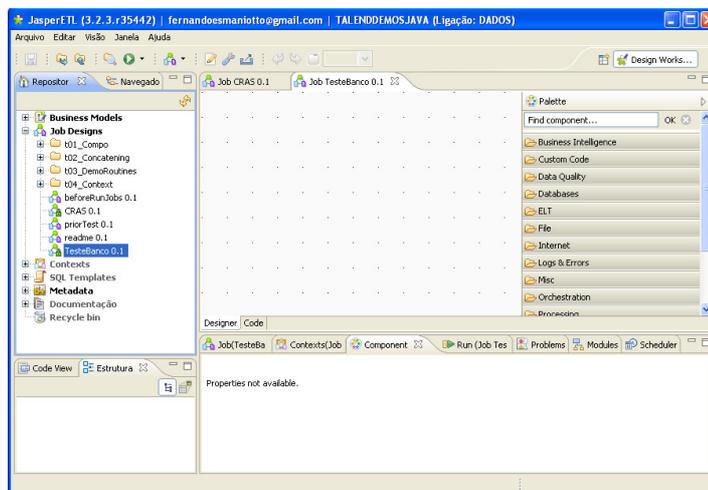


Figura 74 – Job criado
Fonte: Autoria própria.

Feito o procedimento de configuração do Job. Basta incluir agora um objeto para **conexão com o banco de dados**. Para realizar este procedimento será utilizado o componente tMySQLInput_1, este componente é usado para realizar a conexão com banco de dados MySQL. Porém existe a necessidade de configurá-lo.

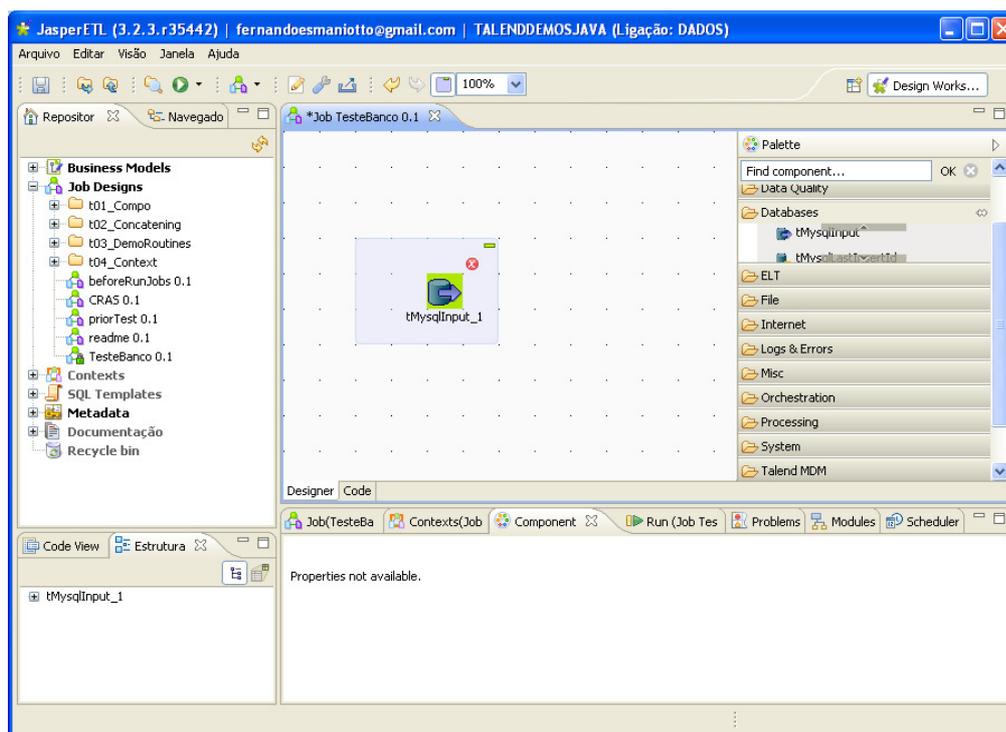


Figura 75 – Inserção do componente tMysqlInput_1
Fonte: Autoria própria.

Para ter acesso ao Assistente de conexão com banco de dados, basta clicar sobre o componente tMysqlInput_1. Na opção Componente existe um botão  usado para configuração de base.

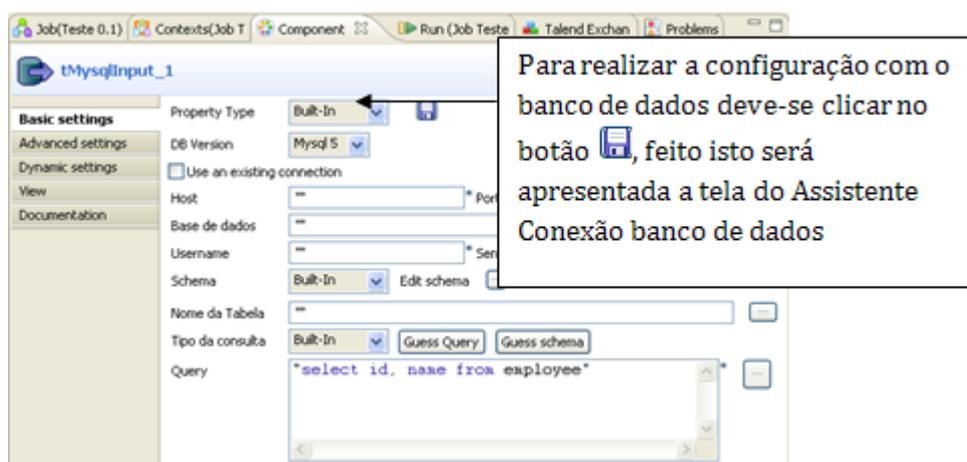


Figura 76 – Opção para realizar configuração com banco de dados
Fonte: Autoria própria.

Logo então, é apresentada a tela de **Conexão com banco de dados**.

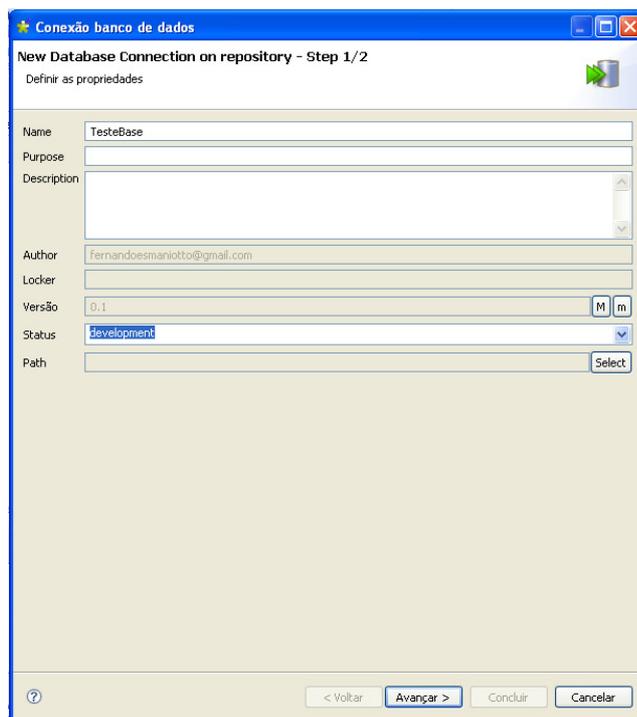


Figura 77 – Configuração do banco de dados
Fonte: Autoria própria.

O **Assistente de Conexão banco de dados** é dividido em duas etapas na primeira deve-se configurar um nome para a conexão. Também é possível colocar uma descrição para fins de documentação. Feito estas configurações clica-se no botão **Avançar**. Na segunda etapa são configurados seis elementos:

1. **Tipo do banco:** Selecionar qual o tipo de banco de dados, para a nova conexão.
2. **Versão do banco:** Selecionar qual a versão do banco de dados.
3. **Usuário:** define o nome de usuário do banco de dados.
4. **Senha:** define a senha do banco de dados;
5. **Servidor:** define o servidor que pode ser local ou remoto.
6. **Banco de dados:** define o nome da base de dados.

Obs.: Tanto o **Usuário** quanto a **Senha** são configurados no momento da instalação do banco de dados no computador.

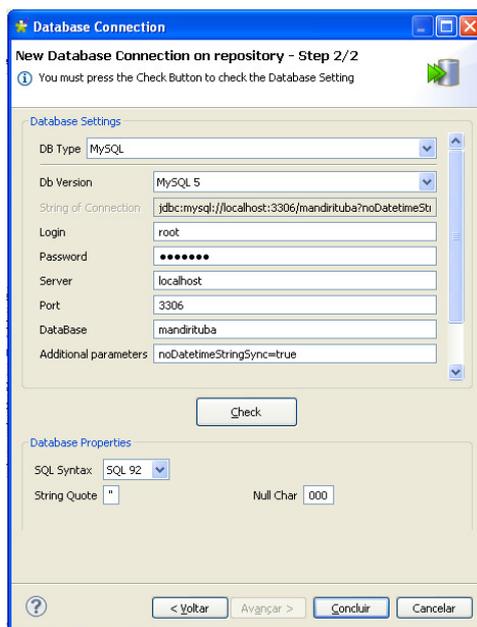


Figura 78 – Configuração das informações sobre o banco de dados
Fonte: Autoria própria.

A figura 78 mostra as configurações que devem ser efetuadas na tela **Conexão com banco de dados**. Após realizar os procedimentos, basta clicar no botão **Concluir**. Antes disso, é interessante clicar no botão **Verificar** para certificar que a conexão está funcionando.

Apesar da conexão com o banco de dados estar completa, ainda é necessário configurar uma **Consulta** para busca de informações na base dados. Para realizar este procedimento basta acessar a aba **Componente** localizada logo abaixo da área de trabalho.

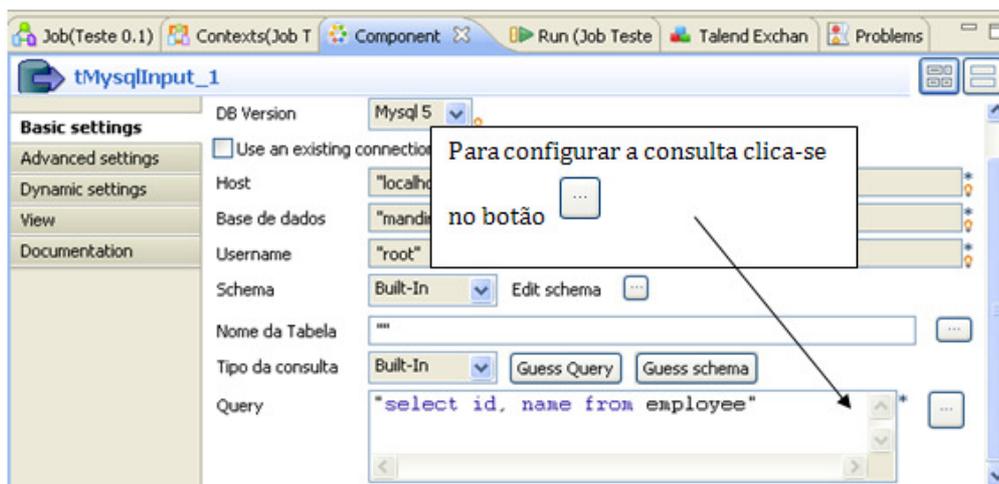


Figura 79 – Botão para configura Consulta
Fonte: Autoria própria.

Clicando no botão  será apresentada a janela **SQL Builder**, nesta parte são configurados três procedimentos.

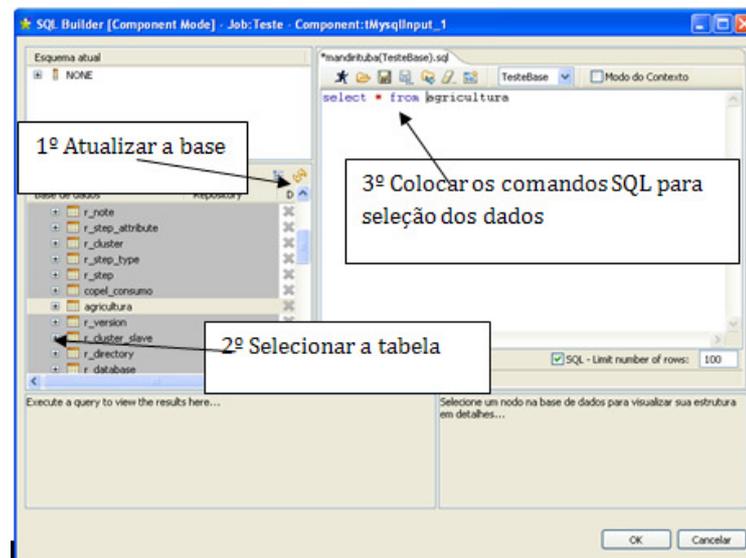


Figura 80 – Opção para realizar a configuração da consulta
Fonte: Autoria própria.

Com os procedimentos finalizados, basta clicar no botão OK. Neste momento é necessário inserir o componente de saída. O componente **tFileOutputExcel_1**, é encontrado na Paleta opção **Arquivo / tFileOutputExcel_1**.

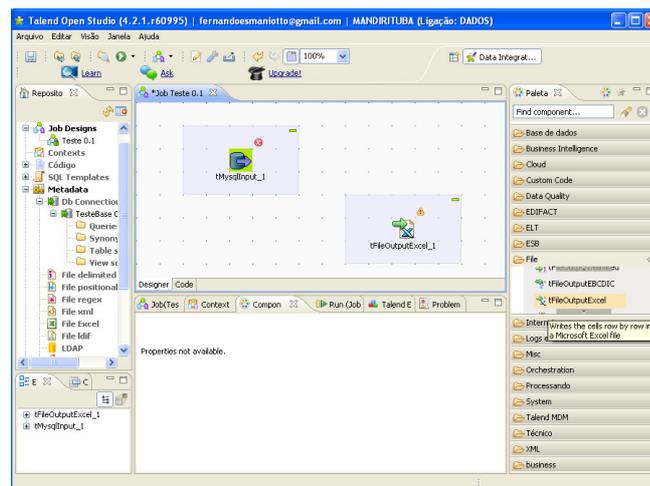


Figura 81 – Configuração da Consulta
Fonte: Autoria própria.

No caso do componente **TFileOutputExcel_1** não existe a necessidade de realizar tantas configurações, apenas é definido o local onde vai ser salvo o arquivo.

Para isso basta clicar sobre o componente e na aba **Componente** localizado abaixo da área de trabalho, tem-se a opção **Nome do arquivo**. Basta colocar caminho do local onde será salvo o arquivo.

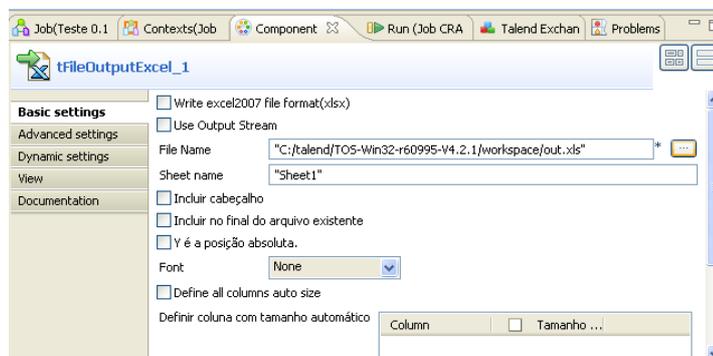


Figura 82 – Configuração do componente tFileOutputExcel_1
Fonte: Autoria própria.

Realizado os procedimentos têm-se os componentes de entrada e saída devidamente configurados. Porém falta ligá-los, para isso basta clicar com o botão direito do mouse sobre o componente tFileMySQLInput_1 e arrastá-lo sobre o componente tFileExcelOutput_. Logo então é possível executar o projeto.

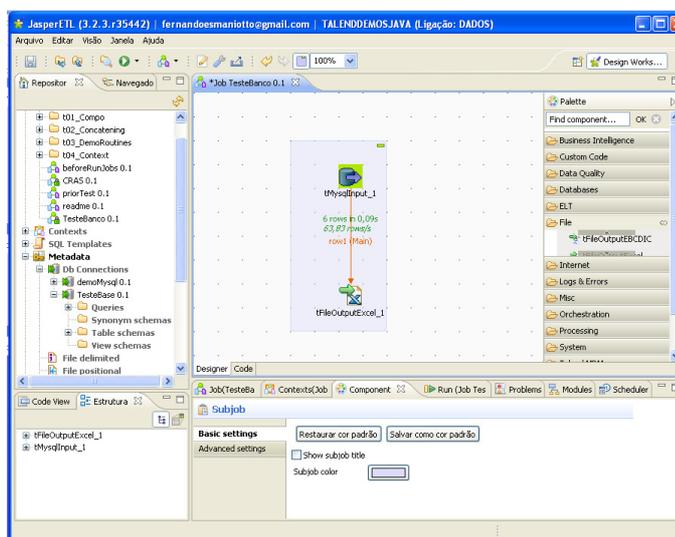


Figura 83 – Ligação entra o banco de dados e o arquivo de saída
Fonte: Autoria própria.

APÊNDICE J – Implementando cubos com a ferramenta Schema Workbench

“O Schema Workbench é um editor de propriedades que facilita a criação de Schemas” (PENTAHO ANALYSIS SERVICES, 2011).

Para criar um Schema é necessário ter uma fonte de dados configurada. Com a fonte de dados configurada, basta acessar o Menu: **Arquivo / Novo / Schema**.

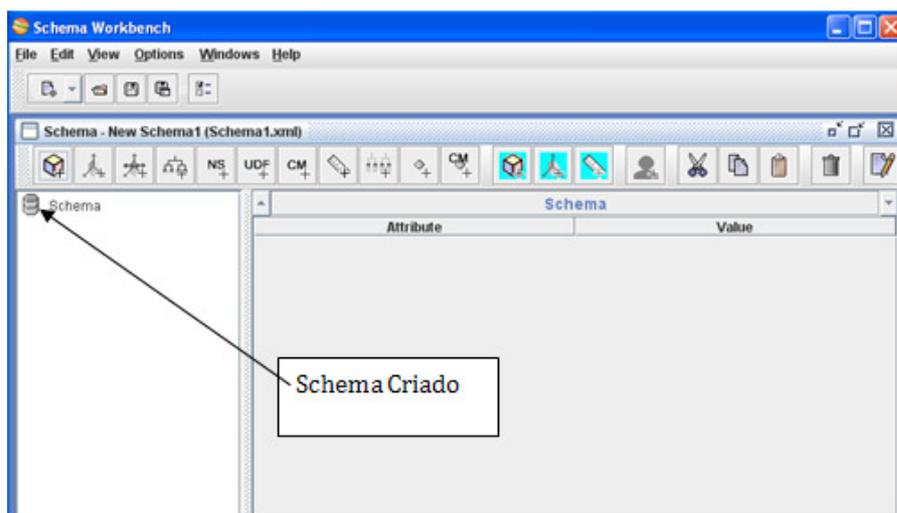


Figura 84 – Implementação do Schema.
Fonte: Autoria própria.

Com o Schema criado é necessário criar um cubo. Para isso, clica-se com o botão direito do mouse sobre o ícone do Schema que acabou de ser criado, ou no ícone . Caso opte pela opção de clicar com o botão direito do mouse sobre o Schema escolha a opção Add cube (Adicionar Cubo).

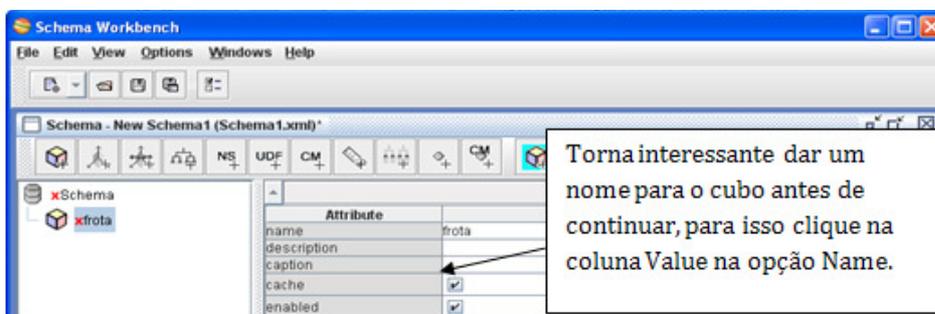


Figura 85 – Implementação do cubo
Fonte: Autoria própria.

O próximo passo é adicionar ao cubo uma tabela. Clicando com o botão direito sobre o cubo escolher a opção Add table (Adicionar tabela). Após adicionar a tabela é necessário escolher com qual tabela o cubo irá trabalhar.

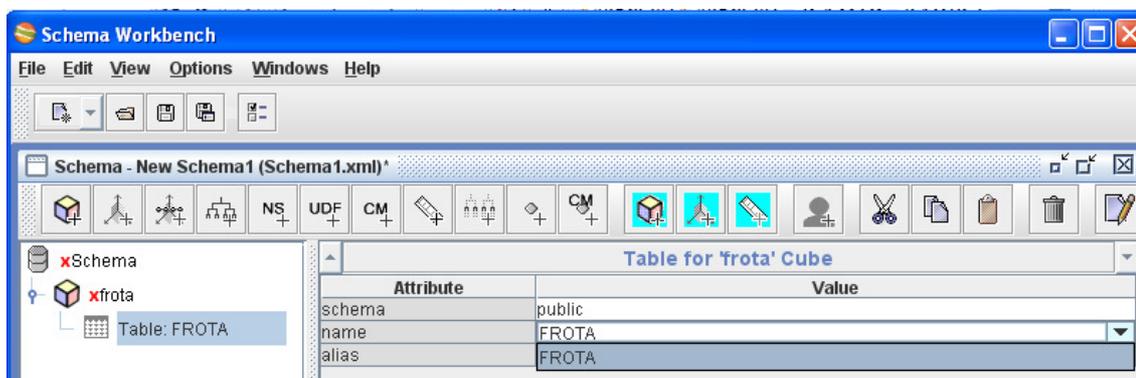


Figura 86 – Escolha da tabela
Fonte: Autoria própria.

Realizado o processo de escolha da tabela é necessário definir regras. Para tal adiciona-se o componente Measure (Medida). Clicando com o botão direito sobre o ícone do cubo e escolhendo a opção Add measure (Adicionar Medida).

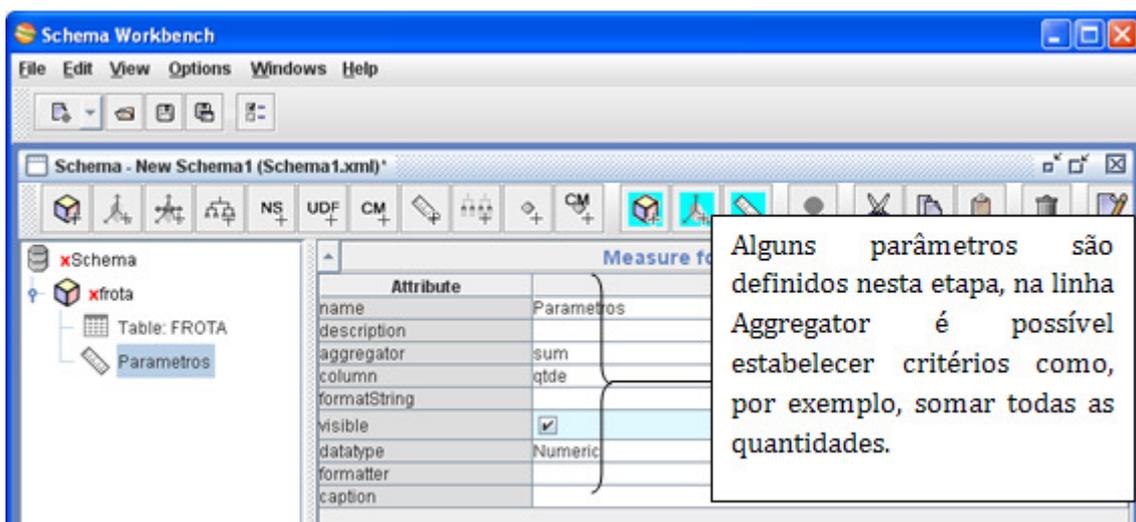


Figura 87 – Definição de parâmetros
Fonte: Autoria própria.

Para finalizar o projeto de implementação do cubo é necessário incluir dois componentes: Dimension (Dimensão) e Hierarchy (Hierarquia). Para isso, deve-se clicar com o botão direito do mouse sobre o cubo e escolher a opção Add dimension

(Adicionar dimensão). Clicando sobre o ícone New dimension (Nova dimensão) escolher a opção New hierarchy (Nova hierarquia).

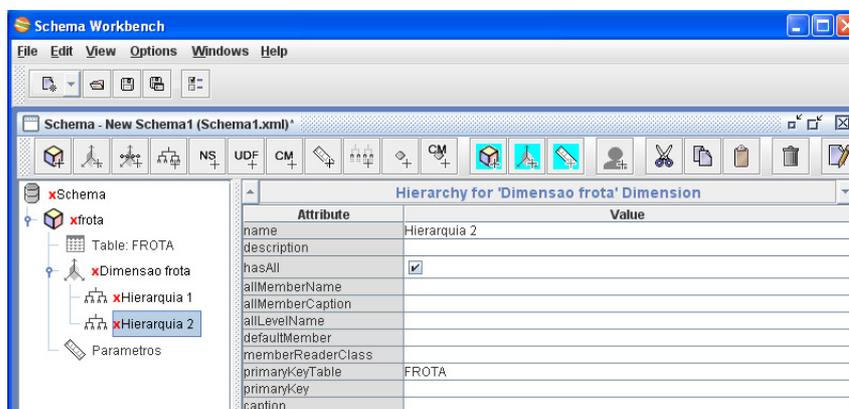


Figura 88 – Adicionando dimensões e hierarquias
Fonte: Autoria própria.

Apesar de adicionar os respectivos componentes, o cubo ainda não está finalizado. Existe a necessidade de vincular as tabelas da base de dados com as hierarquias e também adicionar níveis a elas. Para realizar tal procedimento, basta clicar com o botão direito do mouse sobre a **hierarquia** e escolher a opção Add table (Adicionar tabela). Após adicionar a tabela deve-se então adicionar um nível. Para isso basta clicar com o botão direito do mouse sobre a **hierarquia** e escolher a opção Add level (Adicionar nível). Vale lembrar que esta opção só é habilitada no momento em que se vincula uma **tabela a hierarquia**.

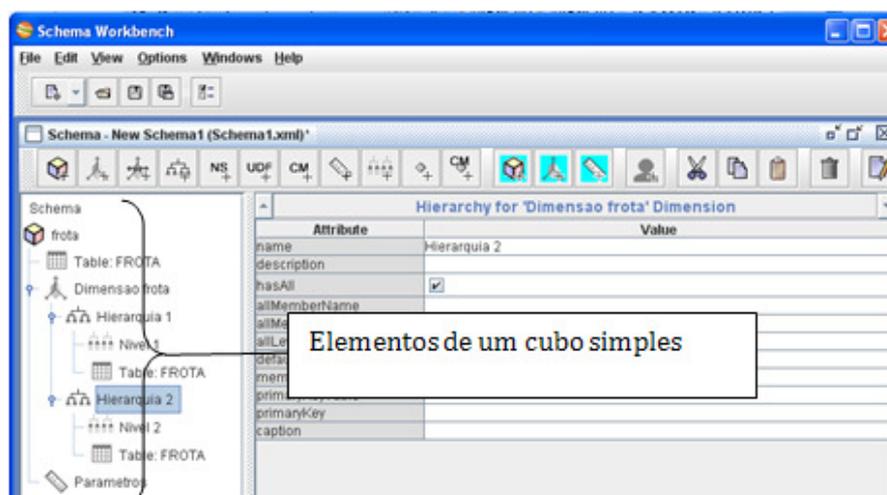


Figura 89 – Elementos de um cubo
Fonte: Autoria própria.

Antes de publicar o esquema é necessário realizar a exemplificação do componente **Tabela** e **Nível** que foram adicionados a **Hierarquia**.

Level for 'Hierarquia 1' Hierarchy	
Attribute	Value
name	Nivel 1
description	
table	FROTA
column	qtde
nameColumn	
parentColumn	
nullParentValue	
ordinalColumn	
type	
uniqueMembers	<input type="checkbox"/>
levelType	
hideMemberIf	
approxRowCount	
caption	
captionColumn	
formatter	

Basicamente dar-se um nome ao nível escolhe a tabela e a coluna a ser utilizada.

Figura 90 – Descrição do componente nível
Fonte: Autoria própria.

A exemplificação do componente Tabela.

Table for 'Hierarquia 1' Hierarchy	
Attribute	Value
schema	public
name	FROTA
alias	

Basicamente escolhe-se a tabela a ser utilizada e o Schema.

Figura 91 – Descrição do componente table
Fonte: Autoria própria.

Falta apenas realizar a publicação do Schema para utilizá-lo no Pentaho. Para realizar este processo deve-se antes modificar o arquivo publisher_config.xml colocando uma senha para que se possa publicar o Schema criado. O arquivo publisher_config.xml é encontrado no diretório: **biserver-ce/Pentaho-solutions/system/**. Após localizar o arquivo, basta abrir e editar colocando uma senha.

```
<publisher-config>
  <publisher-password>senha</publisher-password>
</publisher-config>
```

Figura 92 – Configuração da senha.
Fonte: Autoria própria.

Esta senha vai ser utilizada para realizar o processo de publicação do schema criado. Salve o arquivo e clique no menu **Novo / Publicar**

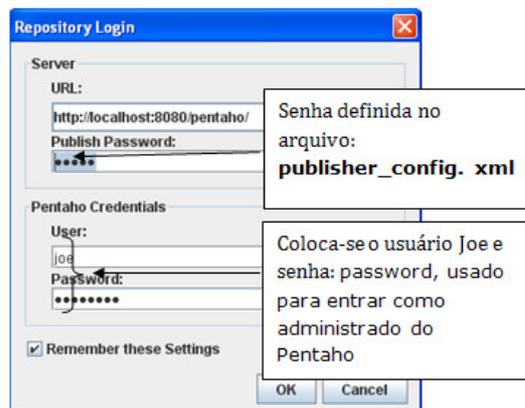


Figura 93 – Definição de acesso para publicação
Fonte: Autoria própria.

Vale lembrar que o servidor do Pentaho deve estar ativo no momento da publicação.

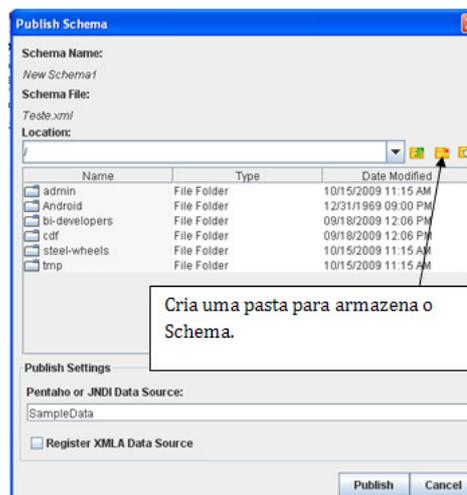


Figura 94 – Tela de publicação do schema
Fonte: Autoria própria.

Nesta etapa basta escolher a pasta e clicando no botão **Publicar** é finalizado o processo de publicação.

APÊNDICE K – Implementando relatórios com a ferramenta Pentaho Report Designer

O Pentaho Report Designer faz parte da suíte Pentaho é uma ferramenta utilizada para geração de relatórios e gráficos. Para criar um relatório é necessário realizar a conexão com a fonte de dados. Com a fonte de dados configurada é possível iniciar o processo para geração de relatório para isso é necessário acessar o menu **Arquivo / Report Wizard** será apresentado à tela do Report Designer Wizard.

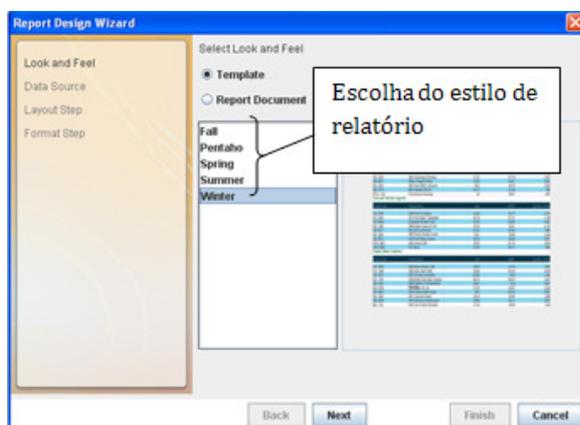


Figura 95 – Tela para escolha de layout para o relatório
Fonte: Autoria própria.

Após escolher o estilo clica-se no botão **Próximo**.

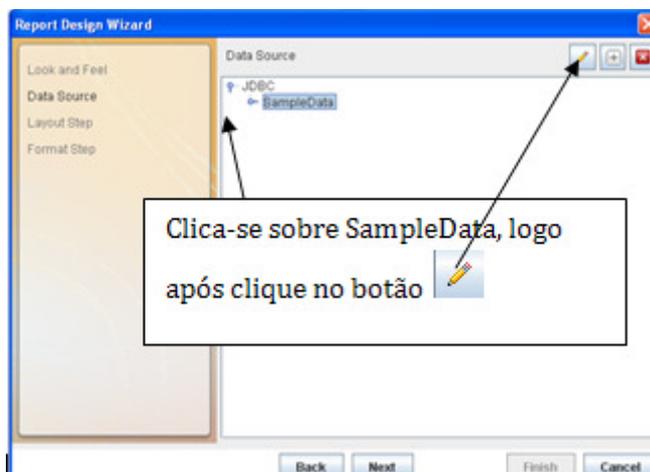


Figura 96 – Tela para escolha de conexão
Fonte: Autoria própria.

Após clicar no botão  será aberta a tela para configuração de conexão.

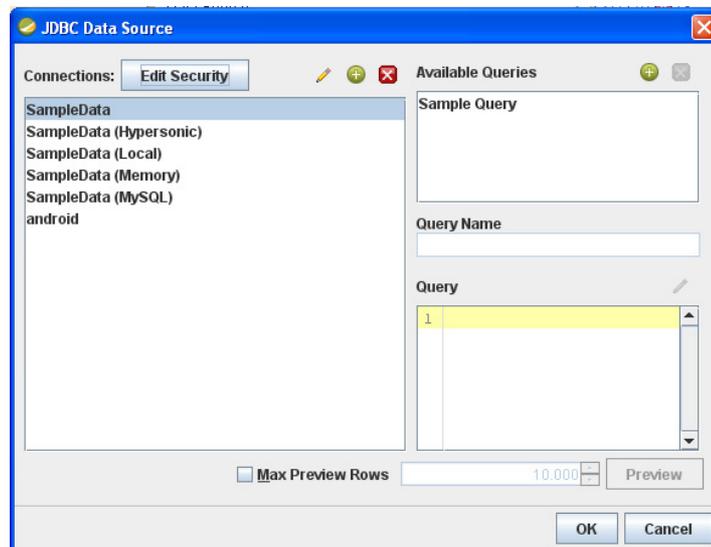


Figura 97 – Tela de exclusão, edição ou alteração de conexão
Fonte: Autoria própria.

Como não existe nenhuma conexão configurada, deve-se clicar no botão  para ter acesso à tela **Conexão com banco de dados**.

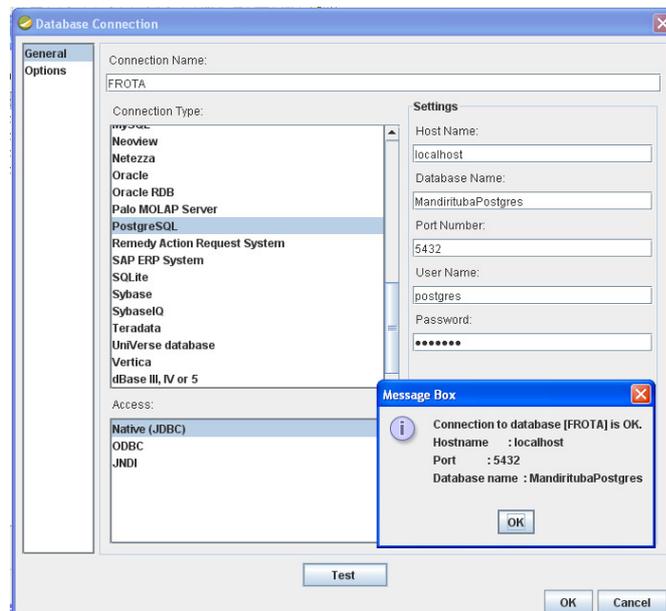


Figura 98 – Tela para realizar a configuração da conexão
Fonte: Autoria própria.

Basta clicar no botão OK para retornar a tela JDBC Data Source. Nesta etapa são inseridos comandos SQL para visualizar os respectivos dados da tabela.

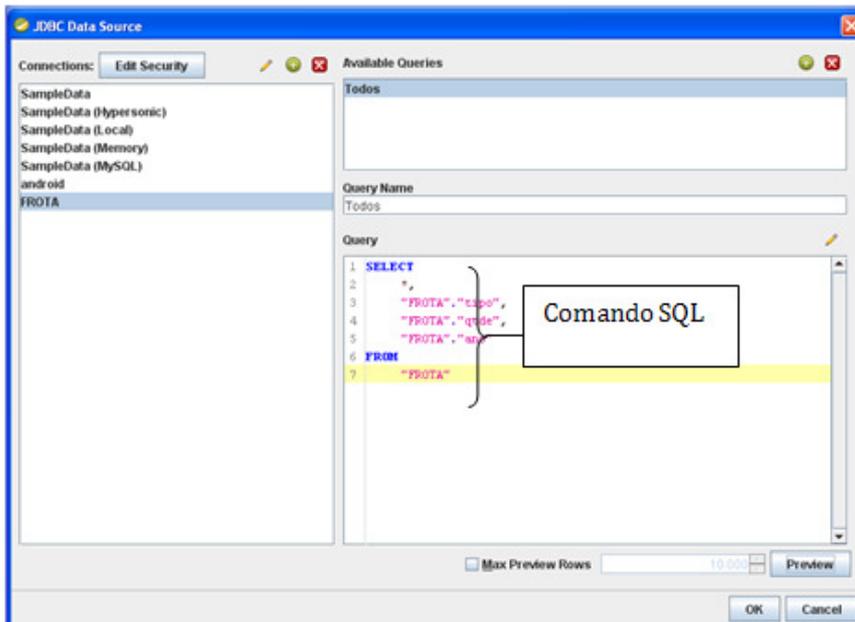


Figura 99 – Tela de inserção de comando para consulta
Fonte: Autoria própria.

Clicando no botão Ok o aplicativo retorna para a tela do Report Designer Wizard, basta no botão **Próximo** para prosseguir. Na próxima tela são escolhidos quais dados serão mostrados no relatório.

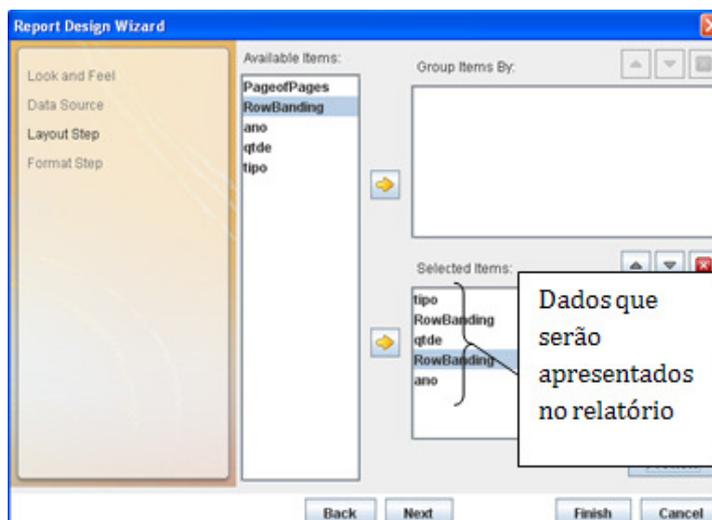


Figura 100 – Seleção dos dados
Fonte: Autoria própria.

Clicando no botão **Visualização** é possível ter uma noção da apresentação do relatório.

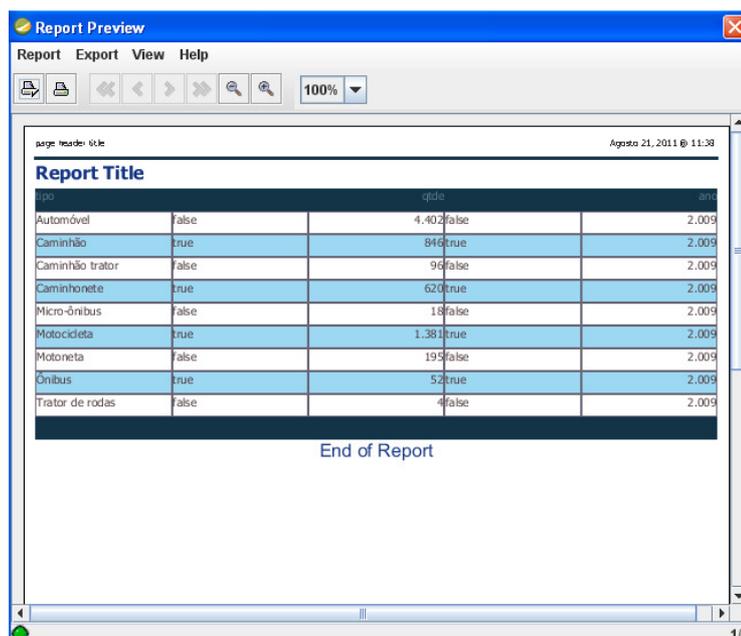


Figura 101 – Visualização do relatório
Fonte: Autoria própria.

Realizado as configurações, basta clicar no botão **Finalizar** e o relatório estará pronto. Porém ainda é possível fazer vários ajustes de acordo com a necessidade.

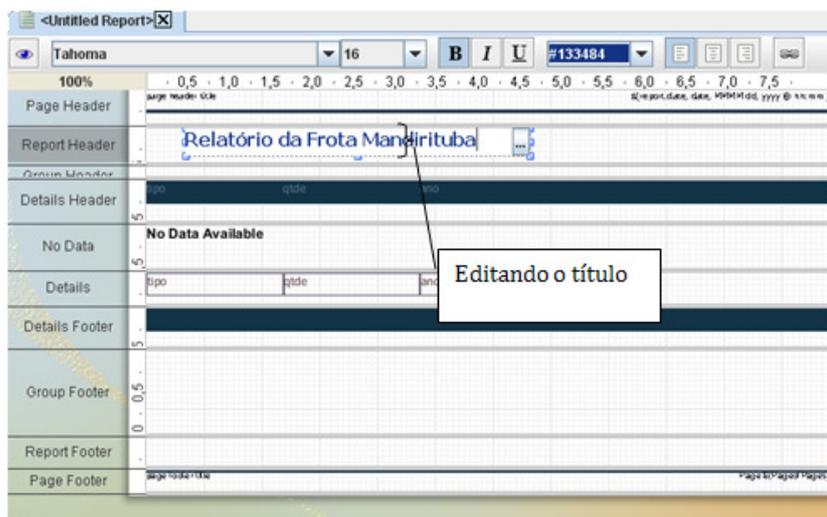


Figura 102 – Modificando o relatório
Fonte: Autoria própria.

Para visualizar as alterações basta clicar no botão  e escolher a opção **Visualização da impressão**.

page header: title Agosto 21, 2011 @ 11:45

Relatório da Frota Mandirituba

Tipo do Veículo	qtde	ano
Automóvel	4.402	2.009
Caminhão	846	2.009
Caminhão trator	96	2.009
Caminhonete	620	2.009
Micro-ônibus	18	2.009
Motocicleta	1.381	2.009
Motoneta	195	2.009
Ônibus	52	2.009
Trator de rodas	4	2.009

Figura 103 – Visualização do relatório alterado
Fonte: Autoria própria.

A figura 103 mostra o mesmo relatório com algumas alterações com relação ao modelo anterior. A geração de relatórios usando o Pentaho Report Designer foi ótima oferecendo total suporte para implementação de um relatório com visual atrativo.