



Análise e Avaliação de Plataformas de Business Intelligence na Nuvem e uma Abordagem da ferramenta Pentaho e suas Integrações .

ANA CRISTINA MEDEIROS

dezembro de 2018



Análise e Avaliação de Plataformas de Business Intelligence na Nuvem e uma Abordagem da ferramenta Pentaho e suas Integrações .

ANA CRISTINA MEDEIROS

Dezembro de 2018

Porto, outubro 2018

Dedicatória

Dedico todo meu esforço e determinação a toda minha família e principalmente ao meu filho.

Resumo

O crescimento exponencial das informações em todos os setores do mercado globalizado está levando as empresas a procurar tecnologias mais sofisticadas, porém com um custo menor para atender suas necessidades. As empresas dependem fortemente de ferramentas mais inteligentes e eficazes de colaboração que possam superar todos os desafios dessa explosão da informação, e a medida que a tecnologia da informação avança surge a necessidade do aumento da complexidade dos sistemas computacionais, a necessidade de novos hardwares que por sua vez requerem novos softwares, com mais flexibilidade e mais facilidade de utilização.

Com toda esta demanda de recursos e inovação, os analistas e empresas estão constantemente a pesquisar novos modelos de tecnologia de negócio e neste âmbito surgiu a proposta da combinação de Computação na Nuvem e Business Intelligence com o objetivo de prover todos os recursos para uma infraestrutura otimizada de BI.

Estas soluções permitem uma otimização eficaz na área de BI, reduz o custo na sua totalidade e proporciona às empresas mais produtividade e competitividade, podendo alterar significativamente a forma de como algumas empresas são geridas. Por exemplo, a terceirização de serviços para provedores de Nuvem elástica ricos em recursos como processamento de dados, número de usuários, armazenamento, etc.

Entender as especificidades da organização é fundamental para que se faça uma escolha mais assertiva de uma solução de BI na Nuvem. Só uma escolha correta capaz de responder sua cultura, pessoas, tamanho, infraestrutura e necessidades da área de negócio permitirá o processo de transformação dos dados em ganho potencializando a vantagem competitiva empresarial.

Sendo este trabalho de natureza exploratória, o método utilizado teve como base a pesquisa aplicada da literatura especializada e a proposta inicial desta dissertação foi um resumo das plataformas de BI na Nuvem mais indicadas segundo o Instituto de pesquisas Gartner, utilizando seus critérios para uma análise comparativa entre essas ferramentas. Também foi aplicado os modelos da ISO/IEC 14598-4 e ISO/IEC 9126-1, para melhor auxiliar os utilizadores de negócio a seleção de uma plataforma mais adequada às necessidades de cada organização.

Tendo em conta a princípio o cenário a ser implementado, foi escolhido uma ferramenta de código aberto segundo as necessidades do negócio, porém por motivos maiores, não foi possível continuar seguindo o percurso planejado desde o início deste projeto. Mas foi possível apresentar uma abordagem das ferramentas escolhidas, a Suite da Pentaho, tendo como objetivo compreender se uma plataforma de licença aberta pode atender a necessidades credíveis de implementação no domínio de BI.

Palavras-chave : Business Intelligence, Computação na Nuvem, Ferramentas de BI na Nuvem, Plataformas de BI licença livre, Suite Pentaho.

Abstract

The exponential growth of information in all sectors of the globalized market is leading companies to look for more sophisticated technologies, but at a lower cost to meet their needs. Companies rely heavily on smarter and more effective collaboration tools that can overcome all the challenges of this information explosion, and as information technology advances, there is a need to increase the complexity of computing systems, the need for new require new software, with more flexibility and ease of use.

With all this demand for resources and innovation, analysts and companies are constantly researching new models of business technology and in this scope the proposal of the combination of Cloud Computing and Business Intelligence came up with the objective of providing all resources for an optimized infrastructure of BI.

These solutions enable effective BI optimization, reduce total cost, and provide enterprises with more productivity and competitiveness, and can significantly change the way some companies are managed. For example, outsourcing services to resource-rich elastic cloud providers such as data processing, number of users, storage, etc.

Understanding the specificities of the organization is critical to making a more assertive choice of a BI solution in the Cloud. Only a correct choice capable of responding to your culture, people, size, infrastructure, and business needs will enable the process of data transformation to gain, leveraging the competitive business advantage.

This work was exploratory in nature, based on the applied research of the specialized literature and the initial proposal of this dissertation was a summary of the BI platforms in the Cloud, according to the Gartner Research Institute, using its criteria for a comparative analysis between these tools. The models of ISO / IEC 14598-4 and ISO / IEC 9126-1 have also been applied, to better assist business users in selecting a platform more suited to the needs of each organization.

Having in mind the scenario to be implemented, an open source tool was chosen according to the needs of the business, but for greater reasons, it was not possible to continue following the course planned since the beginning of this project. But it was possible to present an approach to the chosen tools, the Pentaho Suite, to understand whether an open license platform can meet credible implementation needs in the BI domain.

Keywords: Business Intelligence, Cloud Computing, BI Tools, Business Cloud BI Platforms, Open Source BI Platforms, Pentaho Suite, BI Tools Open source.

Conteúdo

1	<i>Introdução</i>	1
1.1.	Contexto	2
1.2.	Problema	2
1.3.	Objetivos	3
1.4.	Valor da Investigação	3
1.5.	Estrutura do documento	4
2	<i>Estado da Arte</i>	5
2.1	Análise de Valor	5
2.2	New Concept Development Model (NCD)	5
2.3	Identificação de oportunidades	6
2.3.1	Análise de oportunidade	7
2.3.2	Geração e enriquecimento de ideias	8
2.3.3	Seleção de ideias	8
2.4	Valor para o Cliente	9
2.4.1	Valor	9
2.4.2	Análise e Criação de Conteúdo	10
2.4.3	Valor do Cliente	10
2.4.4	Valor percebido	11
2.4.5	Custo da Transação	12
2.4.6	Valor total para o Cliente	13
2.5	Proposta de Valor	13
2.6	Modelo de negócio Canvas	14
3	<i>Contexto</i>	17
3.1	Computação na Nuvem	17
3.1.1	Modelo de Implantação	17
3.1.2	Modelo de Serviço	19
3.2	BI na Nuvem	19
3.2.1	Caraterísticas e Benefícios	21
4	<i>Plataformas de BI na Nuvem</i>	23
4.1	Plataformas Open Source	23
4.1.1	Pentaho	24
4.1.2	Domo	25
4.2	Plataformas Comerciais	26
4.2.1	Sisense	26
4.2.2	Tableau	28

4.2.3	Microsoft Power BI	29
5	<i>Avaliação da solução BI</i>	31
5.1	Avaliação das Plataformas de BI	31
5.1.1	Metodologia Aplicada	32
5.1.2	Os 15 critérios de uma plataforma de BI.....	32
5.2	Análise Comparativa das plataformas de BI.....	34
6	<i>Solução Proposta</i>	39
6.1	Introdução	39
6.2	Âmbito do projeto	39
6.3	Análise do Negócio	40
6.4	Definição dos Requisitos	40
6.4.2	Utilizadores e stakeholders	45
6.5	Definição do Modelo Dimensional	46
6.5.1	Metodologia Kimball para modelação de um DW	46
6.5.2	Data Mart na área Gestão Académica	46
6.5.3	Data Mart na área Gestão de Vendas.	49
6.6	Visão da Implantação – Arquitetura Pentaho BI	51
6.6.1	Definição das tecnologias de integração com Pentaho BI	52
6.7	Visão da Implementação com Pentaho.....	53
6.7.1	Implementação do Data Mart de Vendas no processo ETL.....	53
6.8	Análises dos Dados em Cubos Olap.....	68
6.8.1	Estrutura dos Cubos.....	68
6.8.2	Tabelas agregadas.....	69
6.8.3	Role Playing Dimension e a Dimensão Tempo.....	72
6.8.4	Dimensão Degenerada	73
6.8.5	Alerta de Medidas.....	74
6.8.6	Restrição de Acesso a dados.....	76
6.9	BI Server Pentaho	79
6.9.1	Saiku Analytics	79
6.9.2	Views Dashboards.....	85
6.10	Análise da Suite Pentaho	91
7	<i>Conclusão.....</i>	92
7.1	Melhorias na Implementação	93
7.2	Dificuldades Encontradas.....	93
8	<i>Referências</i>	95

Apêndice A	101
ANEXO A - Dados do Ficheiro LEI 2013	106

Lista de Figuras

Figura 1 -Modelos de Implantação na Nuvem (CENESP, p. 2015)	18
Figura 2-Modelos de Serviços na Nuvem (Azure, 2017)	19
Figura 3- Modelo Framework BI baseado na Nuvem (Fengyu Wang, 2008).....	20
Figura 4- Modelo dimensional gestão do Histórico de notas dos alunos	49
Figura 5-Modelo Relacional Motor Inc.	50
Figura 6-Modelo Dimensional Motor Inc.	51
Figura 7-Arquitetura Geral da Implantação Pentaho BI.....	52
Figura 8-Transformação-Carregamento da dimensão Território na Staging Area	54
Figura 9-Transformação-Carregamento da dimensão Cliente Território na Staging Area	55
Figura 10- Tabela stage_territorio e stage_pais_territorio na BD motor-in-star	56
Figura 11-Transformação- Carregamento da dimensão Cliente no Staging Area	57
Figura 12 Tabela Cliente na BD motor-in-star	58
Figura 13-Transformação- Carregamento da dimensão Cliente no Staging Area	59
Figura 14-Tabela tempo na BD motor-in-star	60
Figura 15-Transformação- Carregamento da dimensão Vendedor no Staging Area.....	61
Figura 16-Tabela vendedor na BD motor-in-star.....	63
Figura 17Transformação- Carregamento da tabela de facto.	63
Figura 18-Carregamento da tabela de facto Vendas na BD motor-in-star	65
Figura 19-Execução do job no processo de transformações de território e cliente_territorio..	65
Figura 20-Execução do job no processo de transformações de Vendedor e Cliente.	66
Figura 21-Execução do job no processo da transformação da tabela de facto.....	67
Figura 22-Execução do processo vários jobs para o carregamento no DW.	67
Figura 23- Definição da tabela Agregada a_t_motor_inc 1ª parte.....	70
Figura 24- Definição da tabela Agregada a_t_motor_inc 2ª parte.....	71
Figura 25- Tabela agregada a_t_motor_inc.	72
Figura 26-Configuração da dimensão tempo no conceito de Role Playng Dimension.	73
Figura 27-Configuração do atributo Situação- Definição do Caption Expression.....	74
Figura 28- Configuração de alertas para medida- Percentual do valor Total de Vendas.....	75
Figura 29-Publicação de um cubo feito no Schema Workbench no BI server.....	76

Figura 30-Configuração do utilizador Antonio no Grupo Latam no PUC	77
Figura 31-Configuração de restrição de acesso aos dados para o grupo LATAM.....	78
Figura 32-Consulta no Saiku com Restrição de Acesso ao Grupo LATAM.....	78
Figura 33-Quantidade Vendidas e o Preço Total de Linhas de produtos por Ano- Tabela	80
Figura 34-Qtd. Vendidas e o Preço Total de Linhas de produtos por Ano – Gráfico de Pizza	81
Figura 35-Qtd. Vendidas e o Preço Total de Linhas de produtos por Ano – Gráfico de Barra ...	81
Figura 36-Receita de vendas de Linhas de Produtos por Ano, Situação e País- Tabela.....	82
Figura 37-Receita de vendas de Linhas de Produtos por Ano, Situação e País- Gráfico.....	83
Figura 38- Aplicação Pivot sobre o resultado da operação Dice em três dimensões	83
Figura 39- Aplicação Pivot sobre o resultado da operação Dice em três dimensões– Visão Gráfica	84
Figura 40-Aplicação de Alerta nas células - Tabela.....	85
Figura 41-Aplicação de Alerta nas células - Gráfico.....	85
Figura 42- Montagem de um Cubo no Saiku e o seu resultado da consulta em MDX	86
Figura 43-Parametros de entrada para criação do novo Dashboard no CDE.....	87
Figura 44- Salvar novo Dashboard CDE	87
Figura 45-Configuração do Dashboard do tipo Blueprint	88
Figura 46-Performace de vendas por Fornecedores, Territórios e Vendedores	88
Figura 47-Drillable Dashboard da Receita de vendas da Empresa Motor Inc	89
Figura 48- Relatório de Vendas na seleção do Cliente Mini Gifti Distribution.....	90
Figura 49- Painel Dinâmico de Dashboards Pentaho-Ctools.....	91
Figura 50-Configuração conexão Base de Dados JDBC/JNDI.....	102
Figura 51-Sequencia ilustrativa de como abrir cubo no Saiku	103
Figura 52-Configuração de utilizadores e grupos no BI Server	104
Figura 53-Dados do ficheiro Lei2013.xls	106

Lista de Tabelas

Tabela 1 Valor percebido pelo cliente em contextos industriais.	11
Tabela 2-Detalhes Técnicos Pentaho (PowerBI, 2017)	25
Tabela 3 Detalhes Técnicos Domo (DOMO, 2017)	26
Tabela 4-Detalhes Técnico Sisense.....	27
Tabela 5-Detalhes Técnico Tableau (Tableau, 2017)	28
Tabela 6-Detalhes Técnico Power BI.....	29
Tabela 7 - Análise comparativa das Plataformas de BI	35
Tabela 8-Requisitos não funcionais	41
Tabela 9-RF Autenticação e Restrição de Acesso	42
Tabela 10-RF Termina da Sessão	42
Tabela 11-RF Navegação Drill Down e Roll Up	42
Tabela 12-RF Visualização dos dados em Relatórios	43
Tabela 13-RF Visualização dos dados em Dashboards.....	43
Tabela 14-RF Visualização dos dados em Gráficos e Tabelas.....	43
Tabela 15-RF Agendamento de Execução de Tarefas	44
Tabela 16-RF Análise Olap.....	44
Tabela 17-RF Exportar Dados para tipos de ficheiros diferentes	44
Tabela 18Tabela de Indicadores de Performance	45

Crónimos e Símbolos

Lista de Acrónimos

BI	Business Intelligence
BD	Data Base
CDA	Community Data Access layer
CDE	Community Dashboard Editor
CDF	Community Dashboard Framework
CEOs	Chief Executive Officer
CIOs	Chief Information Officer
CRM	Customer Relation Management
DD	Degenerate Dimension
DM	Data Mart
DW	Data Warehouse
ERP	Enterprise Resource Planning
HDFS	Hadoop Directory File System
IaaS	Infraestrutura as a Service
IOT	Internet Of Things
JNDI	Java Naming and Directory Interface
KPI	Key Performance Indicator
MDX	Multidimensional Expressions
NIST	Intituto Naciona de Padrões de Tecnologias
ODS	Operation Data Store
OLAP	Online Analytical Processing
PaaS	Plataform as a Service
PB	PetaBytes
PDI	Pentaho Data Integration
PME	Pentaho Metadata Editor
PRD	Pentaho Report Designer
PSW	Schema Workbench

PUC	Pentaho User Console
QBE	Query By Example
ROI	Return Of Investment
SA	Staging Area
SaaS	Software as a Service
SCD	Slowly Changing Dimension
SGBD	Sistema de Gerenciamento de Base de Dados
SQL	Structure Query Language
TB	Terabytes
TCO	Total Cost of Ownership
TI	Tecnologia da Informação
VMs	Virtual Machine
VPN	Virtual Private Network

Glossário

Business Intelligence	Sistema de informação de suporte na tomada de decisão de uma área de negócio. Dashboards Apresentação dinâmica da informação. Data Mart Subconjunto de dados, modelado de forma multidimensional orientado a departamentos específicos das organizações
Data Warehouse	Sistema de armazenamento de dados de uma área de negócio. Base de dados utilizada para os processos de análise dos dados
Online Analytical Processing	Processos que permitem manipular e analisar a informação de múltiplas perspectivas
Staging Area	Área de armazenamento onde ocorrem as transformações dos dados. Camada antes do Data Warehouse.
MDX	O MDX é uma extensão da linguagem SQL que a Microsoft utiliza para consultar e rotear o acesso a dados multidimensionais.
Dashboards	Dashboard é um painel visual que apresenta, de maneira centralizada, um conjunto informações (indicadores e suas métricas).

1 Introdução

À medida que os volumes de dados continuam a aumentar e ter influência direta no comportamento das pessoas e automaticamente nas empresas, os CIOs/CEOs das organizações devem continuar a procurar meios mais inteligentes para ganhar vantagem competitiva no mercado, basicamente através de soluções inovadoras. Entretanto os fornecedores de soluções de BI estão se tornando cada vez mais eficazes para apoiar estas rápidas mudanças no mercado. De acordo com o instituto de consultoria de Gartner (Gartner, 2017), em 2021 número de usuários das modernas plataformas de BI e analítica que são diferenciadas por recursos inteligentes de descoberta de dados crescerá ao dobro da taxa daqueles que não são, e oferecerá o dobro do valor comercial. Para implementar uma 'ferramenta de BI' mais flexível, mais rápido e de custo mais baixo, as organizações estão se voltando para opções baseadas na Nuvem, em oposição às opções do BI tradicional (Chang & Alsufyani, 2015).

O conceito de Nuvem tornou-se muito significativo no mundo de Inteligência de Negócios comparado ao BI tradicional com o aumento de oportunidades de implementação e manutenção de baixo custo de uma plataforma de BI na Nuvem para as pequenas e médias empresas. Com os custos muito mais competitivos para a organização, elas podem disponibilizar soluções mais simples, fáceis de usar e mais adequadas às suas necessidades. Podem ser mais autônomas proporcionando mais independência na área de TI. Também tem capacidade na rapidez da implantação e possibilidades de retorno mais rápido, que somam a vários outros benefícios que o BI na Nuvem oferece (Hamburgo & Buksch, 2016).

Neste contexto foi feita uma análise às ferramentas de BI mais competitivas no mercado atual e com a escolha de uma delas, foi implementado um modelo na área de Vendas utilizando semente tecnologias open source que envolve um processo de BI, mais voltado para pequenas e médias empresas. Ao longo do capítulo é apresentada a motivação para a elaboração desta tese, assim como, a forma como esta vai contribuir na resolução do problema encontrado. São também apresentados os objetivos para esta dissertação.

1.1. Contexto

Com o aumento a cada ano de alunos nas instituições educacionais e consequentemente das informações que estão vinculadas a eles, surgem uma necessidade cada vez maior de orientação e monitorização pelos dirigentes das instituições académicas, com o objetivo de garantir programas mais eficientes e adequados para o progresso educacional de todos os alunos e da instituição. Fazer a gestão do histórico académico de alunos não é possível referente a quantidade exorbitante de registos para análise que um aluno pode produzir. As plataformas de BI na Nuvem é uma proposta de solução mais adequada para as organizações que desejam obter melhores desempenhos e qualidade nas informações utilizando os paradigmas principais de ferramentas ETL, análise de dado através de cubos Olap, Relatórios e Dashboards

1.2. Problema

Como é que se usa o tempo a favor para encontrar vantagem competitiva? E como permitir que os utilizadores de negócios possam agir mais rapidamente na tomada de decisão? As soluções de BI tradicionais se tornaram inflexíveis à medida que os dados se evoluem cada vez mais em diferentes ambientes e a alteração de adaptação das necessidades dos utilizadores. Se olharmos para as aplicações de BI tradicionais, elas ainda são inadequadas para projetos em que uma resposta rápida seja necessária. Os utilizadores estão cada vez mais interessados nas novas plataformas por acharem as ferramentas de BI tradicionais muito complicadas de utilizar. Eles querem ter ferramentas mais autónomas, para que eles possam evitar esse atrasos cruciais e agir rapidamente.

As plataformas de BI são uma opção especial para as empresas que desejam uma implementação de BI mais rápida, flexível, mais fácil de manusear e com valores de licença baixo ou zero.

Este universo ainda está em fase de ascensão e as empresas revelam, ainda, algum desconhecimento relativamente a estas plataformas de BI e, consequentemente, do custo benefício do seu desenvolvimento. Entende-se que seja necessário a divulgação destas ferramentas e seus benefícios utilizando contextos diferenciados para a sua aplicabilidade.

A falta de estratégias de TI mais tangível na maioria das empresas, principalmente nas pequenas e médias empresas mostra uma realidade pouco profissional e de retornos financeiros baixos por vários motivos. Os principais motivos são a falta de inovação no processo, a informação escassa e falta de capital para o investimento em uma tecnologia tão complexa. Divulgar as ferramentas de BI na Nuvem mostrando através da implementação dos processos ETL, análise OLAP e disponibilidade da informação de forma clara e fácil, mostrando todo seu potencial que constitui o principal motivo da presente desta dissertação.

1.3. Objetivos

A primeira fase deste projeto teve como objetivo fazer um levantamento e análise em algumas das plataformas BI na Nuvem destacadas pelo Instituto de Gartner no ano de 2017 e fazer uma classificação entre elas com destaque nos melhores preços, desempenho, usabilidade, etc.

A segunda fase tem como finalidade o desenvolvimento de um sistema de Informação, identificando modelos de dados, análise detalhada do negócio e a exploração das principais funcionalidades da plataforma de BI escolhida na primeira fase, a Pentaho Solution.

Para a primeira abordagem no cenário Acadêmico foi feito o levantamento dos requisitos e toda parte da concepção e análise do negócio, juntamente com o seu modelo dimensional. Este cenário se refere ao Departamento de Informática do ISEP e que tem sua fonte de dados em um ficheiro Excel. Para o segundo cenário, não contendo a parte de análise de negócio, e implementado a partir do modelo dimensional, onde foi explorado a parte técnica da solução utilizando as tecnologias relacionadas ao processo Extraction, Transform and Load(ETL), a Análise Olap com suas tabelas e gráficos, e uma pequena introdução para os Relatórios e Dashboards. Este segundo cenário se refere ao motor de vendas Motor Inc, que tem sua fonte de dados de origem numa BD relacional.

Para finalizar o objetivo do projeto teve como princípio efetuar um percurso de consolidação de conhecimento na utilização de uma ferramenta de BI na Nuvem, a partir da resolução dos desafios que surgem a cada etapa do desenvolvimento.

Objetivos Específicos:

- Aprofundar conhecimentos pela implementação dos conceitos de um Data wareHouse;
- Adicionar e consolidar conhecimentos no modelo dimensional extraída da fonte de dados de um ficheiro Excel;
- Disponibilizar informações através de uma análise gráfica, intuitiva e interativa;
- Implementar um projeto de BI desde o processo de ETL (*Extract, Transform and Load*) até um exemplo de *dashboard* com a ferramenta Pentaho, utilizando todos os seus recursos que engloba este processo de BI.

1.4. Valor da Investigação

A solução a identificar, nas organizações onde for aplicada

- Redução do tempo de implementação, implantação e manutenção do sistema e consequentemente custo de mão de obra;
- Redução de mão de obra de alto nível de especialização;
- Redução de insucesso e abandono de sistemas de BI conforme a adaptação ou complexidade do negócio;

- Entrada de novas funcionalidades e seu funcionamento de imediato;
- Facilidade de migração dos dados num contexto integrado;
- Redução de tempo no processo de carregamento ETL;

1.5. Estrutura do documento

A seleção de uma solução de BI na Nuvem deve ser obtida objetivamente com base em bons critérios, que devem ser muito bem analisados e ponderados de acordo com as necessidades da organização para se ter sucesso nos resultados custo benefício. Neste contexto, a estrutura deste documento se divide em: No capítulo 1 faz-se uma Introdução para melhor entender a análise deste trabalho como contexto, problema, objetivos e Análise de Valor.

No capítulo 2 apresenta-se a importância de se ter um produto de qualidade através da análise de alguns elementos para que possa ter saída no mercado consequentemente este capítulo trata da Análise de Valor, abordando a Análise das necessidades dos clientes, valor para o cliente, proposta de valor e modelo de negócio Canvas.

No capítulo 3 descrevem-se alguns dos conceitos relacionado com este trabalho. A importância da compreensão da Computação na Nuvem que descreve aqui seus modelos de implementação e de serviços. Também neste capítulo se descreve o conceito BI na Nuvem, bem como seus riscos e desafios. No capítulo 4 introduzem-se plataformas open source e comerciais que mais foram visionadas em 2017 pelo quadrante Mágico de Gartner. No capítulo 5 faz-se uma análise das ferramentas utilizando os critérios de Gartner e outro critério de avaliação com o resultado da ferramenta indicada na para uma implementação voltada para o custo baixo, de fácil manuseio e ao mesmo tempo mais dinâmica.

No capítulo 6 temos a solução proposta do negócio de gestão acadêmica que inclui toda a parte de análise como o Escopo do projeto, Análise do Negócio, Análise de Requisitos, Análise de definição do Modelo Dimensional. Neste capítulo temos também a Visão da Implantação com a arquitetura completa do Pentaho e seus componentes. A Visão da implementação se compõe pelo desenvolvimento do processo ETL com a ferramenta PDI. Análise Olap com o Schema Workbench/Mondrian e Saiku um pequeno exemplo de Relatório e Dashboard.

Para finalizar no capítulo 7 apresenta-se na conclusão final as Melhorias e Dificuldades encontradas.

2 Estado da Arte

2.1 Análise de Valor

Com o amadurecimento do mercado em tecnologias, as novas compras de plataformas de BI modernas estão se tornando cultura de massa para quem está acompanhando o mercado, pois as organizações que começaram com pequenas implementações e tiveram sucesso, agora buscam expansão com essas plataformas de BI, trazendo-as cada vez mais modernização e transformando-as, e em alguns casos, em empreendimento padrão para suas organizações (Castro, 2016). A capacidade de lidar com grande volume de dados complexos, administrar implantações, promover conteúdos gerados por usuários para fontes corporativas confiáveis, ampliar e incorporar conteúdo analítico e suportar grandes implementações globais, são alguns dos valores que podem trazer para as empresas.

Com tanta inovação o preço de uma implementação de solução BI diminuiu drasticamente, abrindo oportunidades, de entrada no mercado para as pequenas e médias empresas, e na sequência, para as grandes empresas uma oportunidade de se inovar. A implementação mais rápida dá possibilidades sobre o retorno de investimento (ROI) mais rápido. A disponibilidade das plataformas em dispositivos móveis em qualquer lugar e a qualquer hora e a interação em tempo real entre os analistas de negócio possibilita ainda mais a agilidade da tomada de decisão. Todos estes tipos de recursos favorecem a baixa na receita da empresa e maximiza competitividade (Fengyu Wang, 2008).

.

2.2 New Concept Development Model (NCD)

A grande concorrência no mercado, abriu-se um leque para a inovação e especialmente o desenvolvimento de novos produtos que oferecem ótimas oportunidades de criação e vantagens competitivas. (Reguia, 2014)

Alguns autores (Cooper 1988; Cooper & Kleinschmidt, 1991, 1994; Frishammar & Florén, 2008; Herstatt & Verworn, 2001; Langerak et al., 2004) tiveram a conscientização e estudaram sobre a importância das primeiras etapas dos processos de inovação denominado “front end off innovation”, assim conseguiram provar a grande influência que tem sobre a inovação e os sucessos dos produtos.

NCD- New Concept Development Model, apresentado por Peter é um modelo que pertence ao processo de inovação e se divide em 3 fases: Fuzzy Front-End (FFE), New Product Development (NPD) e Comercialização. Este modelo fornece uma linguagem comum e uma visão holística do front-end e que por sua vez se divide em três áreas distintas como: motor, ou centro do modelo. É a parte que dirige os cinco elementos do Front End de Ideias(FEI) e é abastecido pela visão, estratégia e cultura da organização e que impulsiona a inovação. A segunda ou parte interna define os cinco elementos de atividade da frente. O terceiro elemento consiste nos fatores ambientais externos. O FEI, possui a forma em círculo pois indica que as ideias e oportunidades são interligadas, pois, uma boa ocasião para gerar ou testar uma ideia vem da criação de uma oportunidade, assim como uma ideia pode levar a uma oportunidade e também pode-se exigir uma ideia para aproveitar uma oportunidade (Koen, 2001). Este conceito faz com que o modelo possua duas entradas como uma ideia gerada e uma oportunidade identificada.

Os cinco elementos Front-End são: Identificação da oportunidade; Análise da oportunidade; Geração e enriquecimento de ideias; Seleção de ideias e Definição de Conceito.

O tema apresentado nesta dissertação surgiu da necessidade de melhorar a consulta aos históricos dos alunos do instituto informático do ISEP, da qual era feita através de ferramentas não dinamizadas com resultados questionáveis. Identificaram-se uma oportunidade para a implantação de uma ferramenta acessível, ao preço, flexível e dinâmica em análise e painéis de visualização. Logo a seguir vamos utilizar os elementos do NCD que foi aplicado para a análise de valor nas plataformas de BI.

2.3 Identificação de oportunidades

A Business Intelligence evoluiu rapidamente nos últimos anos. Até no início do ano 2.000, o modelo ainda era o do legado tradicional BI - ou seja, a trabalhar com grandes conjuntos de dados, exigindo hardware significativo, e levando muito dinheiro e tempo para implementar e implantar a solução. Cerca de uma década atrás a Computação na Nuvem contribuiu com aceleração do processo de BI. A primeira oferta baseada nesta tecnologia foi lançamento da versão beta da Elastic Compute Cloud da Amazon (EC2) (Amazon, 2017). O Amazon fornecia aos seus clientes via Web Service a condição de adaptar a sua capacidade computacional à capacidade requerida pelas aplicações a suportar. Alguns dos melhores conceitos desta solução é Flexibilidade, elasticidade e pay-per-use. (Gonçalves, 2012)

O BI na Nuvem elimina a necessidade de usar uma mistura de ferramentas, permitindo que os utilizadores construam modelos de dados complexos em um ambiente visual e dando-lhes a liberdade de explorar, filtrar e detalhar esses dados em qualquer granularidade de forma independente e com respostas instantâneas para queries dinâmicas (CRISTESCU, 2016).

Ferramentas de BI permitem, reduzir custos e desperdícios, melhorar a performance do trabalho, do colaborador ou do departamento, identificar oportunidade de negócio e de crescimento com mais eficácia. E a melhora na capacidade de avaliar a relação com clientes e fornecedores dá à empresa mais vantagem competitiva (Castro, 2016).

BI na Nuvem prova que não é preciso um grande investimento e recurso de TI para lidar com dados complexos. Enquanto existem outras plataformas que permitem transferir tarefas de processamento de dados para o departamento de TI, as soluções de BI na Nuvem removem a carga de trabalho em vez de transferi-las entre as partes interessadas (Chang & Alsufyani, 2015)

2.3.1 Análise de oportunidade

As decisões de compra neste segmento de BI estão agora sendo muito influenciadas por utilizadores empresariais que exigem produtos mais fáceis de utilizar, mais baratos e que ofereçam valor comercial claro, que permitam processamentos mais rápidos e análises mais poderosas sem a exigência de especialistas em TI. Estas ferramentas pretendem aumentar o controlo e disponibilidade sobre a informação que estes têm ao seu alcance e, consequentemente, torná-los mais independentes, eficientes e produtivos. Não ter que esperar pelas equipas de IT e todo o fluxo necessariamente inerente, que vai desde a intenção de obter determinada informação até receber o relatório final de acordo com o inicialmente idealizado, permite agilizar os processos de tomada de decisão. Esta ferramenta proporciona às organizações um melhor nível de produtividade devido à redução de tempo de resposta necessário para realizar um relatório, principalmente quando existe a necessidade de cruzar informação de diversos departamentos. Ajuda na flexibilidade pois cada utilizador pode personalizar os seus relatórios, fazer as suas próprias queries e receber as respostas mais rápidas sem recorrer às equipas de IT, sendo cada vez mais relevante consoante o aumento progressivo da quantidade de dados. Estes tipos de ferramentas são escaláveis, podendo ser adaptado para qualquer tipo e tamanho da empresa. Logo abaixo são apresentados alguns benefícios na utilização destas plataformas de BI na Nuvem (Stipié & Bronzin, 2012).

Rápida implantação: agilidade na implementação e implantação, sem necessidade de construir uma infraestrutura própria impulsiona um rápido retorno sobre o investimento (ROI); (Stipié & Bronzin, 2012)

Baixo investimento: ao evitar os custos relacionados a implementação de um BI tradicional, é possível baixar o custo total de propriedade -TCO, consequentemente seu produto/serviço pode ter seu custo reduzido; (Stipié & Bronzin, 2012)

Mobilidade avançada na distribuição dos dados: o acesso remoto aos dados em tempo real facilita a colaboração e tomada de decisão. Qualquer dispositivo móvel pode acessar relatórios, registros de clientes, requisitos do projeto, e mais, elevando a produtividade das equipas;

Escalabilidade: capacidade de suportar um aumento de número de utilizadores simultâneos, em diferentes localizações, com custos mais baixos e menor tempo para se instalar um hardware ou software adicional;

Atualização e manutenção: sem custos extras as empresas podem usufruir das atualizações e manutenções contínuas, ou utilização de equipe própria para este fim;

Multi-tenant: é possível gerir todas as bases de clientes com um único servidor, podendo compartilhar o mesmo ponto de entrada e distribuir custos entre seus clientes o que diminui no custo por utilizador;

Baixo risco: O baixo investimento representa um menor risco para as empresas.

Portabilidade: oferecem plataformas para diferentes tipos de dispositivos;

Computação elástica: é possível variar na demanda do serviço - número e velocidade de processamento, de acordo com a necessidade do serviço;

Armazenamento dinâmico: de acordo com os dados a armazenar oferece funcionalidade com múltiplos tipos de armazenamento. Estes dados são replicados pelos múltiplos servidores, sendo garantida a sua durabilidade, mesmo em caso de falha; (Stipié & Bronzin, 2012)

Segurança de dados: as soluções de BI em Nuvem oferecem uma gama de sofisticados protocolos de segurança, como por exemplo criptografia e segregação de dados, segmentação de rede de apoio, medidas de autenticação, impressão digital de utilizador, entre outros; Computação abstrata: oferece infraestrutura de hardware através de virtualização (Castro, 2016).

2.3.2 Geração e enriquecimento de ideias

Em conjunto com os dois colaboradores do projeto efetuou-se um brainstorming com a forma de apresentar as ideias para solucionar o problema encontrado. Foram realizadas reuniões para discutir as ideias e para se perceber quais as ferramentas que poderiam ser selecionadas e que se aproxima mais do negócio proposto. Entretanto foram evoluiu a ideia de que os dados poderiam ser diretamente extraídos da Base de Dados do instituto, exigindo assim uma escolha de ferramenta mais robustas e quais seriam as técnicas de apresentação mais adequadas para os utilizadores finais.

2.3.3 Seleção de ideias

A ideia num primeiro instante é selecionar uma plataforma de BI na Nuvem para implantar no Departamento de Informática do Isep, sendo assim tivemos que aplicar alguns métodos para que fosse possível a escolha da ferramenta mais adequada que suprimisse as necessidades atuais e futuras da instituição. Um desses métodos aplicados foi a análise de Gartner, que teve uma primeira classificação das plataformas, em seguida foi aplicado alguns critérios qualitativos sobre estas plataformas afim de delimitar o mais possível a leitura relacionada as capacidades críticas mais importantes das ferramentas segundo o cliente.

2.4 Valor para o Cliente

2.4.1 Valor

Ao determinar o critério da Estratégia de Visão para a Oferta (Produto), a Gartner avaliou a capacidade dos fornecedores para suportar as principais tendências que gerarão valor comercial em 2017 e pelos próximos anos. Os produtos e funções existentes e planejados que contribuem para as tendências acima foram tidos em conta na pontuação de cada fornecedor

Infraestrutura: o suporte a um mercado (compradores e vendedores) em que organizações, clientes e parceiros podem comprar e vender aplicativos analíticos personalizados, fontes de dados agregadas, visualizações personalizadas e algoritmos que se integram à plataforma BI e analítica; O acesso nativo a uma variedade de fontes de dados NoSQL como Hadoop, Spark e outras fontes de dados do NoSQL, bancos de dados de gráficos ou bancos de dados de pesquisa como Elastic search e Kibana, Attivio ou Splunk; e Suporte para implantações híbridas em locais e Nuvem.

Gestão de dados: um catálogo de dados ágeis e com curadoria onde os usuários corporativos podem pesquisar, acessar, encontrar e avaliar dados internos certificados, bem como dados externos abertos e premium com fluxo de trabalho; A preparação inteligente de dados em dados multiestruturados é um recurso visionário fundamental nesta categoria; Enriquecer e inferir relacionamentos automaticamente (para gerar automaticamente um modelo para análise) para fazer recomendações para melhorar ou aprimorar informações; A capacidade de promover automaticamente modelos e conteúdo gerados pelo usuário para o SOR e reutilizar e construir sobre variáveis, cálculos, modelos e conteúdo existentes é fundamental para o autoatendimento de confiança de grande escala; o processamento moderno de push-down para grandes fontes de dados, automatizando a seleção de onde processar melhor uma consulta, é um recurso importante para avançar para suportar conjuntos de dados grandes e complexos, alavancando o grande processamento de dados e minimizando a necessidade de mover dados; o apoio à preparação, harmonização e alavancagem de eventos em tempo real e dados em fluxo contínuo.

Partilha de resultados: a capacidade de invocar ações empresariais dentro da plataforma, seja em um painel de bordo ou incorporado em outra aplicação, representa um nível de sofisticação além do suporte mainstream atual para alertas condicionais e desencadeamento de eventos com base em eventos do sistema; O NLG ou a narração de insights é a próxima fase da evolução do relatório padrão para narração de histórias com texto descritivo para aumentar o conteúdo representado visualmente; Recomendações contextuais para conteúdo relevante - com base na visão adquirida com a colaboração e a interação social pelos usuários - substituirão em grande parte a necessidade de compartilhar conteúdo e descobertas manualmente em toda a organização; a simulação integrada de apontar e clicar, a análise e a otimização do que e otimizar estendem os tipos de análise que os usuários precisam, que hoje são geralmente criados usando cálculos personalizados; cada vez mais, as organizações precisam renderizar conteúdo analítico em experiências imersivas para diferentes tipos de usuários em várias telas sensíveis ao toque.

2.4.2 Análise e Criação de Conteúdo

As capacidades de detecção de patentes inteligentes e automatizadas que automatizam a identificação dos padrões e clusters escondidos em dados que muitas vezes são perdidos pelos analistas explorando manualmente conjuntos de dados, são fundamentais para as plataformas de BI e analítica de próxima geração. A identificação automatizada de resultados é a chave para permitir e expandir o acesso a análises para mais usuários dentro da organização e acelerar o tempo de percepção enquanto reduz o viés.

O NLP para voz e texto para apoiar o conceito de um assistente de análise pessoal que pode gerar consultas de idioma natural e explicar suas descobertas para usuários que usam NLG será uma interface futura dominante para análise.

Suporte para uma ampla gama de análise de conteúdo e análise de texto em relação a dados não estruturados à medida que as organizações exploram novas fontes de informação.

2.4.3 Valor do Cliente

Tabela 1 Valor percebido pelo cliente em contextos industriais.

Domínio	Escopo		
	Produto	Serviço	Relação
Benefícios	Facilidade de uso; Capacidade de suportar requisitos de dados complexos; Tipos de complexidade dos utilizadores; Preços atrativos; Produtos escaláveis; Gerenciamento de metadados; Conteúdo analítico incorporado; Administração de plataforma de BI; BI na Nuvem; ETL autônomo e armazenamento de dados; Exploração e criação de dispositivos móveis	Responsividade Flexibilidade Confiabilidade Personalização Paga-por-uso Suporte técnico Inovação -uso	Confiança Solidariedade
Sacrifices	Em contradição aos Benefícios vai depender de cada ferramenta	Interoperabilidade entre Sistemas; Vulnerabilidade na rede e na virtualização; Integração entre sistemas abrangência da banda larga Segurança Dificuldade em Migrações, ampliações do Sistema	Conflitos de Governança de Riscos de Conformidade Normas legais Confidencialidade dos dados;

2.4.4 Valor percebido

Os clientes tendem a fazer suas escolhas avaliando qual oferta proporciona maior valor, ou seja, maximizam o valor, dentro dos limites impostos pelos custos envolvidos na procura e pelas limitações do conhecimento, mobilidade e renda. Eles formam uma expectativa de valor e agem com base nela. A probabilidade de satisfação e repetição da compra depende da oferta atender ou não a essa expectativa de valor (Scharf & Soriano-Sierra, 2008).

Como não podemos analisar o valor percebido em várias plataformas de BI, vamos aplicá-lo então na ferramenta Pentaho, a ferramenta escolhida segundo nossa análise, para a implementação deste projeto.

2.4.4.1 Pentaho

Benefícios

É uma plataforma de código aberto, e sua arquitetura e tecnologia foi desenvolvida e focada na melhoria de funcionalidades ao invés de migração de software legado, que é mudado a cada versão e consequentemente ficando sujeito a muitos erros, por se tratar da adequação a um grande volume de dados e a novos padrões criados para os softwares novos. São riscos grandes que geram custos e tempo, pois além das versões serem demoradas para ser criadas, os softwares depois de implantados demoram para serem estabilizados e muitas vezes quando são, já não satisfazem mais a necessidade de seus clientes. Neste cenário a Pentaho proporciona a montagem rápida das soluções personalizadas de BI, aumentando o valor e a capacidade das soluções de suas empresas, criando propostas superiores as tradicionais do mercado a um custo significativamente baixo (Freitas, 2015).

Oferece uma gama de funcionalidades através da preparação de dados, do autoatendimento e da análise avançada, com foco especial no grande acesso e integração de dados com PDI, PME, PRD, Weka, etc. Abaixo descreve-se alguns benefícios desta ferramenta.

- Usado para uma ampla gama de casos de uso de BI, possui uma forte presença no mercado OEM e embutido de BI;
- Analisa dados tradicionais (SQL, arquivos ad hoc, bancos de dados NoSQL) e não estruturados (como feeds de mídia social, dados de log e fluxos de dados de máquinas de fontes do IoT). Forte em acesso e integração de dados e o custo da licença oferece integração nativa com tecnologias como Hadoop, Spark, Cassandra e MongoDB.
- Funcionalmente classificadas como excelentes: áreas de administração, segurança e arquitetura, ETL autônomo e armazenamento de dados, análises avançadas incorporadas, conteúdo analítico incorporado, conectividade de fonte de dados e gerenciamento de ingestão de metadados, e classificadas como boas: preparação de dados de autoatendimento, exploração visual interativa, painéis analíticos e exploração e criação de dispositivos móveis.
- Foco de suas estratégias de marketing: grande análise de dados e a IoT com a organização Hitachi Insight (Gartner, 2017).

2.4.5 Custo da Transação

Em termos de esforço, uma implantação para pequenas e médias empresa o Pentaho é ferramenta mais indicada. Uma ferramenta de código aberto com uma arquitetura desenvolvida do zero, que utiliza padrões populares e que pode ser facilmente integrada a qualquer processo de negócio e está a todo momento sendo atualizada pela comunidade de desenvolvedores.

Devido ao seu modelo de negócio de código aberto, a Pentaho não cobra por taxa de licença de software, mas possui uma versão comercial para quem necessite de manutenção e suporte, que inclui suporte por telefone e na web. Porém por se tratar de uma plataforma de fácil aprendizagem e muito sugestiva, com um pouco de conhecimento mais técnico dispensa este tipo de gasto. Um módulo (painéis, relatórios, análise, integração de dados ou mineração de dados) custa US \$ 11.000 por ano para até quatro CPUs sem limites de usuários, de acordo com a empresa. (team, 2017-11-23)

2.4.6 Valor total para o Cliente

2.4.6.1 Pentaho

Valor do Produto: posicionado no quadrante Nicho de Mercado mostra grande capacidade em segmentos específicos no mercado. (análises voltadas para o cliente, relatórios ágeis e dashboarding, incorporação ou grande integração de dados). Todos os benefícios citados no item Valor Percebido. Apesar de possuir uma capacidade limitada de inovação e de não superar outros fornecedores em usabilidade, a diferenciação de seu preço e o acesso à integração de dados pontua em grande para a escolha desta ferramenta.

Valor do Serviço: possui recursos de implementação e suporte ou bases de clientes relativamente limitadas para grandes volumes e análise de dados, não possuindo solidificação de posições no mercado. Pode cobrar, dependendo do módulo por manutenção e suporte, que inclui suporte por telefone e na web durante o horário comercial local do cliente;

Valor do Funcionário/Imagem: Pentaho enfrenta um desafio contínuo na área de suporte ao cliente. Possui menos capacidade para suporte para publicações, compartilhamento e colaboração, descoberta de dados inteligentes (falta de geração automática das informações) e consulta de linguagem natural;

Custo monetário para o cliente: Licença GPL para a versão Comunidades e se sentiremos a necessidade de mais recursos, dependendo da lógica do negócio, habilitação do Utilizador ou suporte, existe a versão Comercial, que é paga. O cliente pode optar na contratação de uma empresa especializada para a implementação, que também poderá auxiliar com o treinamento da equipe e com o suporte da ferramenta. Mesmo assim, por ser código aberto, sua requisição representa um baixo custo, pois o cliente não terá despesa de sua aquisição ou licença. A ferramenta na sua versão Comunitária cobre uma grande gama de necessidades de vários tipos de negócios, uma alternativa totalmente viável para qualquer negócio que não pode ter altos custos com ferramentas de gestão.

2.5 Proposta de Valor

"O PODER DOS DADOS EM SUAS MÃOS"

Com ferramentas extremamente intuitivas e interativas que podem integrar múltiplas fontes, permitindo análises mais inteligentes, geração de relatórios e dashboard personalizados, para

ajudar na tomada de decisão. Com facilidade, agilidade e segurança, em qualquer lugar e a qualquer hora. É um produto que explora plenamente enormes quantidades de dados e tem recuperação precisa, conhecimento de alta qualidade, mais precisamente para pequenas e médias organizações e até mesmo para profissionais que trabalham individualmente, melhorando a produtividade e impulsionando a competitividade de sua empresa. Oferecer o poder dos dados ao alcance das mãos dos utilizadores de negócio com as soluções revolucionárias de 'Cloud BI'.

Entregar valor ao negócio é isto. É ser ágil, é entregar no tempo e no custo correto e ter a possibilidade de responder rapidamente às mudanças.

2.6 Modelo de negócio Canvas

Utilizando um modelo de Design Thinking, Alexander Osterwalder, desenvolveu o modelo Canvas que se difere dos outros modelos de negócio por incentivar a Inovação, prototipação e cocriação. O excelente aspecto do Modelo de Negócio Canvas permite que você visualize o “encaixe estratégico”, baseado em nove componentes básicos como: Segmentos dos Clientes; Proposta de Valor; Canais; Relacionamento com Clientes; Fontes de Receita; Recursos Principais; Atividades Chave; Parcerias Principais; Estrutura de Custo. E suas relações, que demonstram de forma lógica como uma organização é capaz de gerar valor. O nosso modelo não cobre todos os componentes, porém cobre as principais áreas que se pode ver a seguir.

Parceiros Chaves

- Empresas que fornecem serviços na Nuvem;
- Engenheiros de Sistema

Atividades Chaves

- Fornecimento de Dados, processos e análises centralizados e compartilhados;
- Serviços por demanda (BD, Softwares, Infra, Plataformas, Dados, Analise);
- Gestão Multi-tenant: servidor central com gestão de vários clientes. Disponibilização das Informações atualizadas para a análise através de Ferramentas interativas Dashboards e Relatórios;
- Facilidade de interação entre os Gerentes para tomada de decisão.

Proposta de Valor

- Oferece produtos e serviço sobre demanda. Mais agilidade e interatividade nos serviços entre utilizadores finais, fornecedores e Gestores;
- Gastos generalizados minimizados (implantação manutenção, elasticidade, escalabilidade);

- Elasticidade nos dados (flexibilidade na disponibilização da banda larga, tempo de processamento, espaço de memória) pago-por-utilização;
- Diminuição na latência (tempo de requisição/resposta) para tomada de decisão
- Ganho de espaço físico;
- Computação abstrata;
- Mobilidade Avançada (Acesso aos dados através de dispositivos em qualquer lugar);
- Escalabilidade (integração de outros softwares/dispositivos/etc. quando necessário);
- Portabilidade (pode ser acessado através de diferentes plataformas);
- Suporte e atualizações automáticas;
- Rápida implantação: Pode impulsionar um rápido retorno sobre o investimento (ROI);
- Melhoria da produtividade e colaboração a análises preditivas;

Relação com os clientes

- Fornecedor dos (SaaS, IaaS e PaaS) / Rede Privada, Rede pública ou híbrida;
- Suporte/Manutenção online;
- Atendimento pós-venda;

Segmento de mercado

- Plataformas podem ser utilizadas em qualquer segmento

Recursos Chaves – Ativos da Empresa

- Análise de Artigos (Jornal, Conferências, Estado da Arte)
- Infraestrutura;
- Integração de outras ferramentas, APIs, OEM, BI embutido;
Software; Servidor de dados

Canais:

- Artigos; Tese; Dissertação;
- Habilidade do utilizador (Comunidades/Webinars/gerenciamento automático de informações);
Redes sociais.

Estrutura de Custo:

- Licença do Produto ou Serviço (relativo)
- Manutenção (relativo)

3 Contexto

3.1 Computação na Nuvem

De acordo com o Instituto Nacional de Padrões de Tecnologia (NIST), "A computação em Nuvem é um modelo para habilitar o acesso por rede ubíquo, conveniente e sob demanda a um conjunto compartilhado de vários recursos computacionais configuráveis (como redes, servidores, aplicações armazenamentos e serviços) que possam ser rapidamente provisionados e liberados com o mínimo de esforço de gerenciamento ou interação com o provedor de serviços. (Stipié & Bronzin, 2012).

Computação na Nuvem é uma tendência recente de tecnologia que surgiu da necessidade de construir infraestruturas de TI complexas, com a realização de instalação, configurações e atualizações de sistemas de hardware e software. Seu objetivo é proporcionar serviços de Tecnologia da Informação (TI) sob demanda com pagamento proporcional à utilização de recursos ou dados do utilizador (Sousa, et al., 2011).

3.1.1 Modelo de Implantação

Quanto ao acesso e a disponibilidade há diferentes tipos de modelos de implantação para os ambientes de computação na Nuvem. Nem todos os utilizadores podem aceder a todos ou determinados recursos no ambiente de trabalho, por isso o acesso aos serviços em Nuvem dependem do processo de negócio, do tipo de informação a ser disponível e do nível de visão desejado. (Sousa, et al., 2011).

Nuvem privada: esta infraestrutura de Nuvem se utiliza de uma rede inteligente e flexível que provê uma experiência de uso confiável, possibilitando o armazenamento e o acesso à informações e dados corporativos com segurança. Utilizada e administrada pela própria empresa (pode ser administrada por terceiros), local ou remotamente. As técnicas de gerenciamento de redes, configurações dos provedores, utilização de tecnologias de autenticação e autorização são empregados na política de acesso. (Sousa, et al., 2011).

Este modelo é a opção certa para empresas que atuam em segmentos altamente regulamentados ou que trabalham com informações confidenciais pois fornece mais estabilidade para o armazenamento dos dados corporativos na Nuvem, garantindo controle

total sobre o ambiente com menor risco de ameaças vindas de terceiros. Porém, o preço de implantação de uma Nuvem interna pode ser um empecilho para algumas empresas de pequeno e médio porte. Além disso, o custo de operação contínua da Nuvem pode exceder o custo de uso de uma Nuvem pública.

Nuvem pública: Este modelo representa uma infraestrutura de propriedade pública, administrado e operado por um provedor de serviços especializados, por exemplo como uma organização governamental, uma instituição acadêmica, ou uma combinação entre eles e disponibilizada para o público em geral sendo acessado por qualquer utilizador que conheça a localização do serviço sem restrições de acessos. (Sousa, et al., 2011).

É um modelo que tem como um de seus benefícios a redução de custos e, portanto, é uma boa alternativa para empresas que tenham um orçamento restrito ou outras prioridades.

E é indicada para empresas de pequeno e médio porte e que trabalham com dados menos sigilosos.

Nuvem comunitária: utilizada por uma determinada comunidade, ou seja, um grupo de pessoas de diferentes organizações que partilham dos mesmos objetivos e interesses (por exemplo, missão, requisitos de segurança, política e considerações de conformidade). Pode ser de propriedade, gerenciada e operada por uma ou mais organizações da comunidade, por um terceiro, ou por alguma combinação deles, e pode existir dentro ou fora das instalações.

Nuvem híbrida: esta infraestrutura se conecta por meios de tecnologias proprietárias ou padronizadas com uma ou mais tipos de nuvens (privada, pública ou comunidade) e que permite a comunicação de dados e a portabilidade de aplicações (como transferência de processamento para a Nuvem para balanceamento de carga entre nuvens). (Stipié & Bronzin, 2012). Por exemplo, sistemas críticos ou que manipulam informações confidenciais podem ser hospedados internamente enquanto outros sistemas, que não lidam com dados sigilosos, podem ser utilizados em uma rede pública.

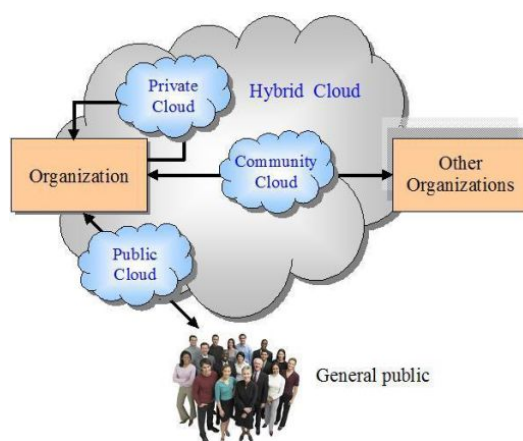


Figura 1 -Modelos de Implantação na Nuvem (CENESP, p. 2015)

3.1.2 Modelo de Serviço

O ambiente de computação na Nuvem define um padrão arquitetural para este tipo de solução e é composto de três modelos de serviços, oficialmente reconhecido pelo NIT. (Sousa, et al., 2011).

Infraestrutura como serviço (IaaS): O principal objetivo do IaaS é prover uma infraestrutura para o PaaS e o SaaS tornar mais fácil e acessível o fornecimento de recursos. Este modelo oferece sistema operacional, redes, servidores, armazenamento, e aplicativos, onde usualmente os utilizadores tem controle sobre estes recursos e podem selecionar componentes de redes. Porém eles não controlam e nem administra a infraestrutura na Nuvem. (Reis & Silva, 2014). O Amazon Elastic Cloud Computing (EC2) (Reis & Silva, 2014) e Eucalyptus (Hill, 2007) são exemplos de IaaS.

Plataforma como serviço (PaaS): Aqui o modelo oferece plataformas de integração para implementar e testar as aplicações na Nuvem. Os recursos são sistema operacional, linguagens de programação e ambientes de desenvolvimento para as aplicações. O desenvolvimento e publicação de suas aplicações e possivelmente, suas configurações ficam por conta do utilizador. Google App Engine (Ciurana, 2009) e Microsoft Azure (Azure, 2016) são exemplos de PaaS.

Software como um Serviço (SaaS): O fornecedor é responsável por toda infraestrutura da aplicação e pela aplicação em si. Ele disponibiliza o sistema de software através de uma interface cliente (VM) como um navegador web. O cliente paga pelo uso da aplicação sem se preocupar com o backend. A grande vantagem é que o software pode ser acessado a qualquer momento, permitindo maior integração entre outras unidades de das empresas ou serviços. Ex. Customer Relationship Management (CRM) da Salesforce e o Google Docs.

A figura 2 mostra uma visão generalizada dos modelos de serviços.



Figura 2-Modelos de Serviços na Nuvem (Azure, 2017)

3.2 BI na Nuvem

Segundo (Costa & Santos, 2009), pode se dizer que os sistemas de Business Intelligence combinam a recolha de dados operacionais, permitem o seu posterior armazenamento em

repositórios adequados, que por sua vez vão permitir a gestão de conhecimento através de diferentes ferramentas de análise, exploração e apresentação da informação, essencial para a tomada de decisão. Os sistemas de Business Intelligence têm como objetivo principal melhorar a disponibilidade e qualidade da informação (Costa & Santos, 2009). Para isso, estes sistemas incluem Data WareHouse ou Data Marts, que são repositórios de informação organizacional, incluem também sistemas de processamento analítico de dados (On-Line Analytical Processing - OLAP) e mecanismos de Data Mining (DM). Em termos de análise, a ferramenta OLAP proporciona uma análise multidimensional dos dados e o Data Mining são utilizados para a extração de conhecimento, identificando padrões e modelos preditivos nos dados.

As estratégias das empresas que trabalham com BI estão sendo reavaliadas em consequência do crescimento exponencial e contínuo do volume de dados. A maioria delas está enfatizando uma implementação mais flexível e mais adaptável ao meio externo e o uso de sistemas estratégicos em tempo real. Business Intelligence com base na Computação Nuvem que também pode ser definido como 'Cloud BI' é um conjunto de serviços que permite aos utilizadores aceder e trabalhar com dados corporativos, com mais facilidade, sem ser especialista em TI, dando a possibilidade de criar as suas próprias pesquisas e análises. Também com base na computação abstrata fica mais flexível e, mais rápido a disponibilização de recursos sobre medida (redes, armazenamentos, servidores, aplicativos e serviços) para as empresas que querem otimizar seus gastos. Portanto, a escalabilidade dos dados e seu aprimoramento com regulamentação personalizadas dão ao segmento de BI uma ampla gama de oportunidades (Alsufyani & Chang, 2015).

Na figura 3 mostra-se um exemplo de um framework de BI baseado na Nuvem subdividido em 7 camadas. (Fengyu Wang, 2008)

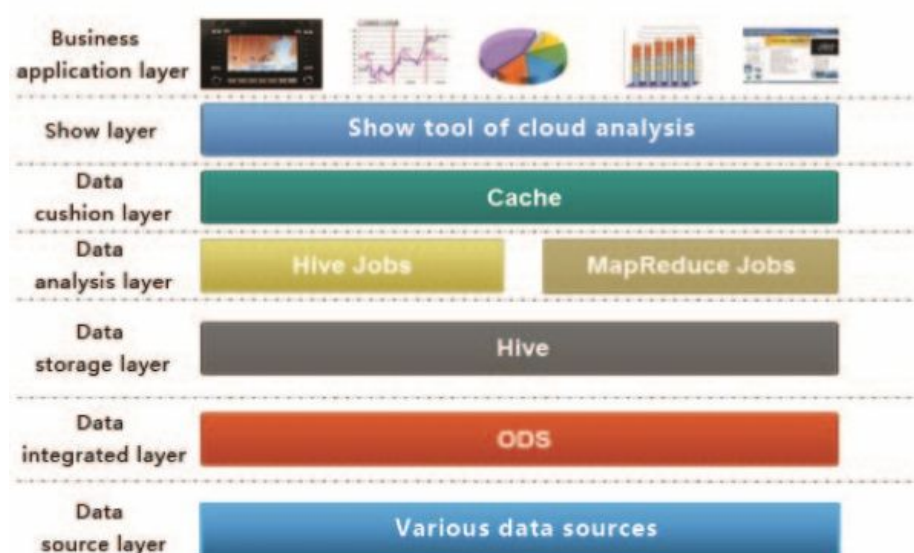


Figura 3- Modelo Framework BI baseado na Nuvem (Fengyu Wang, 2008)

Camada da fonte de dados: – A primeira fase de todo o processo de BI começa na coleção de informação que seja potencialmente necessária para todo o processo de decisão. Contém componentes que vão recolher a informação das várias fontes (i.e.: recursos humanos, finanças, faturação, transacionais, CRM, ERP). Os dados podem ser estruturados (numa base de dados ou sistema de estrutura de dados) ou não estruturados (documentos, imagens, e-mails), podendo ser obtidos tanto de fontes internas como externas. (Castro, 2016).

Camada de integração de dados: Faz a integração, manipulação e transformação de dados heterogêneos vindos de diversas fontes para formar a visão unificada conectada dos dados, podendo-se dividir em diferentes tópicos de acordo com as áreas do negócio. Um ODS destina-se a relatórios operacionais e suporta requisitos de relatórios atuais ou em tempo real, enquanto que um data wareHouse é destinado a relatórios de análise histórica e de tendências geralmente em um grande volume de dados. (Fengyu Wang, 2008).

Camada de armazenamento de dados e camada de análise de dados: Os dados estruturados no ODS são armazenados em forma de textos no HDFS suportado pelo Hadoop. Facilita a compactação de dados, consultas ad-hoc, e a análise de conjuntos de dados extensos. A camada de análise com MapReduce permite melhorar a capacidade de computação para reduzir o tempo de processamento dos dados; (Fengyu Wang, 2008).

Camada de Lógica: Permite selecionar pontos de vistas diferentes através de análises de sumários de dados e construção de resumos de uma ampla gama de dados simplificando a complexidade lógica do front end e melhora a velocidade no tempo de acesso aos dados;

Camada de Apresentação: Aqui a ferramenta desenvolvida é independente da camada de negócio e permite a utilização de diversos métodos de análise. Data Mining e KDD;

Camada de aplicação de negócios: Ferramentas ou tecnologia que permite a apresentação dos dados, a maior parte sobre forma de relatórios, gráficos, infográficos, painéis de controle(dashboards), etc.

3.2.1 Caraterísticas e Benefícios

Segue abaixo uma lista dos principais benefícios que o BI na Nuvem oferece:

Rápida implementação: Sem a necessidade de construir uma infraestrutura própria, a agilidade na implementação pode impulsionar um rápido retorno sobre o investimento (ROI);

Baixo investimento: Ao evitar os custos associados com o desenvolvimento e implementação de uma estrutura de alojamento tradicional, assim será possível baixar o TCO, custo total de propriedade e reduzir o custo de seu produto/solução;

Mobilidade avançada na distribuição dos dados: Este benefício permite que os profissionais tenham acesso remoto aos dados, em tempo real, facilitando a colaboração e tomada de decisão. O acesso de relatórios, registos de clientes, requisitos do projeto, e-mails, a partir de qualquer dispositivo móvel eleva a produtividade das equipes;

Escalabilidade: A tecnologia de Cloud Computing suporta qualquer aumento no número de utilizadores simultâneos, independente da sua localização, sem o custo ou o tempo necessário para implantar e instalar um hardware ou software adicional;

Upgrades e manutenção: A manutenção dos sistemas de informação pode ser automatizada. Isso permite uma disponibilidade de dados mais rápida que alimenta os aplicativos de BI, e diminuição no custo com este serviço.

Multi-tenant: Com uma única instalação de servidor (conceito de multi-tenant), é possível gerir todas as bases de clientes, compartilhando o mesmo ponto de entrada e distribuindo os custos entre eles, gerando assim um menor custo por utilizador.

Computação elástica: Com picos de utilização variados para dar resposta às mudanças, geralmente controlada com ferramentas de monitorização de sistemas, a computação elástica possui a capacidade de expandir ou diminuir rapidamente recursos informáticos de processamento, memória e armazenamento. Assim a empresa tem uma diminuição no custo por evitar a pagar por recursos não utilizados ou inativos. (Azure, 2017)

Armazenamento dinâmico: Funcionalidade oferecida com múltiplos tipos de armazenamento de acordo com os dados. Estes dados são replicados pelos múltiplos servidores, sendo garantida a sua durabilidade, mesmo em caso de falha.

Segurança de dados: As soluções de BI em Nuvem oferecem sofisticados protocolos de segurança, como por exemplo criptografia e segregação de dados, segmentação de rede de apoio, medidas de autenticação, impressão digital de utilizador, entre outros.

Pay-per-use: Pago pelo que é utilizado, dispensa o utilizador da obrigação de reservar por uma quantidade exata de recursos. Ou seja, se for necessário, ele tem a possibilidade de aumentar a quantidade de recursos reservados sem grandes esforços e de forma automática (GONÇALVES, 2012).

4 Plataformas de BI na Nuvem

Sempre foi um grande desafio para as empresas fazer com que os projetos de Business Intelligence (BI) tivessem menos desgastes emocional, de tempo ou económico, devido sua complexidade. Esse tipo de solução sempre exigiu, profissionais muito bem qualificados com conhecimentos específicos em uma variedade de domínio na área de TI, e custo alto em todo processo que o envolve (Castro, 2016).

A implantação de uma solução de BI na Nuvem tem um interesse especial para as organizações que desejam melhorar principalmente na agilidade e redução de custos, e explora todo os benefícios que as plataformas possam oferecer.

Hoje existem inúmeras plataformas de BI na Nuvem no mercado a trabalhar a unificação dos dados, o processamento dos mesmos para a análise e visualização.

Neste capítulo apresenta-se uma análise que foi realizada entre 7 plataformas de BI na Nuvem. Estas ferramentas foram escolhidas entre as 24 selecionadas no quadrante Mágico do Grupo de Gartner. Este Grupo é altamente reconhecido no mercado mundial pelas suas pesquisas e consultoria pela qualidade de suas informações na elaboração de relatórios no mundo de TI.

É importante lembrar que as empresas utilizam o quadrante como opção no momento da escolha da ferramenta que mais se aproxima a sua realidade, por isso não necessariamente uma solução classificada como líder atende a todos os requisitos de uma corporação. Tudo depende da necessidade específica da empresa.

Segue abaixo a descrição das plataformas que nós escolhemos para analisarmos. Esta seleção teve como base as necessidades do caso de uso aqui relacionado.

4.1 Plataformas Open Source

4.1.1 Pentaho

Uma plataforma de código aberto desenvolvido em Java para soluções empresariais. Possui sua versão GPL e que pode ser executada em qualquer plataforma que tenha uma JVM padrão. A solução de BI com Pentaho se compõe de uma suíte de negócios que se divide em duas partes: Pentaho BI Plataforma, implementada na forma do servidor web Tomcat, e a segunda cliente de desenvolvimento, que criam conteúdo para a plataforma. Para estes ambientes estão disponíveis componentes para execução de processos ETL, que fazem carga de Data Warehouse, criação de relatórios pré-formatados e ad hoc, cubos OLAP, painéis (Dashboards) e mineração de dados (Data Mining). Esses recursos podem ser combinados para soluções mais sofisticadas. A plataforma de BI executa todas as soluções de BI como serviço, podendo prover acesso a soluções externas via serviços web Soap. (Pentaho, 2016).

A plataforma Web se divide em:

O Solution Engine e seus componentes que através de uma máquina de execução de workflow interna executa e sequencia as chamadas de cada componente para o resultado desejado.

O Portal, permite que o cliente final navegue entre as soluções e acione a execução de qualquer recurso, como um relatório ou dashboard.

Alguns serviços como registro de soluções, controle de acesso, relatórios ad-hoc, agendamentos são pré-configurados no BI Server. Permite ainda que alguns novos serviços sejam criados e implementados livremente.

Os principais módulos do Pentaho BI são:

Pentaho Report Designer (PRD), que conecta a qualquer fonte de dados para qual exista um driver **JDBC**. Relatórios pixel, resultado de fórmulas, sub-relatórios, links, imagens, gráficos (pizza, barra, linha etc.) são recursos deste módulo;

O Pentaho Metadata Editor (PME), permite ao usuário de BI criar visões de negócios independentes através de agrupamento de campos de tabelas que tenham alguma correlação entre elas;

Pentaho Schema Workbench (PSW), possui uma interface Visual que permite navegar entre as definições do cubo para criar métricas, dimensões e hierarquias (Devmedia, 2016)

Pentaho Design Studio (PDS), é o ambiente de implementação de Soluções de BI que cria Servlets e as combinam em soluções mais complexas;

O **Pentaho Weka** é um ambiente gráfico para Data Mining. O usuário cria e testa hipóteses através dos dados (Devmedia, 2016).

O software de integração de dados e processos ETL, **Pentaho Data Integration (PDI)**, permite uma mistura de informações vindas de várias fontes incluindo Hadoop, bases de dados relacionais e bases de dados analíticos. Oracle, PostgreSQL, SQLServer, são alguns formatos SGBD que ele é capaz de ler e escrever, além de importar arquivos texto (csv ou fixo), planilhas Excel e base de dados ODBC. Possui uma gama de bibliotecas, análise de informações em tempo real, controle total com painéis de arrastar e soltar interativos e personalizáveis baseados na

Web e de visualizações completas que permitem que os utilizadores vejam dados com clareza, ampliem informações e vários detalhes importantes. Sua integração com plataformas móveis, permite aos utilizadores desfrutar-se das interfaces suaves com seus smartphones. (Devmedia, 2016)

4.1.1.1 Detalhes Técnicos

Tabela 2-Detalhes Técnicos Pentaho (PowerBI, 2017)

Dispositivos Suportados	Suporte de línguas	Modelo de Pagamento	
Windows /Android iPhone/iPad Mac	USA, UK, Canada, Austrália, Alemanha India, Latim América Oriente Médio	Free e Baseado em Cota	
Implantação	Tipos de Clientes	Satisfação dos Usuários	Pontuação
SaaS	Small Business Large Enterprises Medium Business	95%	8.1

Clientes notáveis: Vodafone, MySQL, Motorola, Globosat, FDA, Caixa Econômica, Banco do Brasil.

4.1.2 Domo

É a primeira ferramenta de Business Intelligence (BI) de autoatendimento open source. Tem um sistema baseado em Nuvem confiável e oferece um conjunto de ferramentas de dados muito amplo e suporte de conector entre as ferramentas testadas e um conjunto exclusivo de recursos de colaboração social. Pertencente a um ecossistema aberto. Permite que qualquer desenvolvedor crie e entregue aplicativos para o Appstore do Domo. O Business Cloud do Domo possui aplicativos nativos do Android e do iOS que fornecem insights para você tomar decisões onde quer que você esteja (DOMO, 2017).

Como o Domo é usado principalmente por pessoas de negócios para painéis de estilo de gerenciamento, e muitas vezes é implantado em linhas de negócios com pouco ou nenhum suporte de TI, uma percentagem maior de seu relatório de referências de clientes usando-o principalmente para casos de uso descentralizado

O Domo oferece também uma equipe de especialistas da Universidade DOMO para resolver seus problemas, e se preferirem por conta própria utilizar o banco de dados do DOMO University que cheio de relatórios de suporte, vídeos de testemunhos, whitepapers e aulas interativas que podem ensinar você a trabalhar com ele.

Sua presença geográfica inicial e em expansão e uma visão do produto que se concentra em fechar lacunas atuais que o classifica no Quadrante de Gartner Nicho de mercado no eixo Abrangência da Visão (Gartner, 2017).

Domo oferece três níveis:

- Starter (para equipes pequenas)
- Profissional (usuários aumentados, dados e controle administrativo)
- Enterprise (sem limites, controles de nível empresarial)

Domo está entre os Tops 10 para os 3 tipos de softwares como BI, DM e o Data Analysis.

4.1.2.1 Detalhes Técnicos

Tabela 3 Detalhes Técnicos Domo (DOMO, 2017)

Dispositivos Suportados	Suporte de Línguas	Modelo de Pagamento	
Windows Android iPhone/iPad Baseado na Web	USA, UK, Canadá Internacional	Gratuito, baseado em Cota Pág. Mensal	
Implantação	Tipo de cliente	Satisfação Usuário	Pontuação
Hosp. Nuvem Hosp. Local API código Aberto	Pequenas e Médias Empresas	94%	9.1

Cientes Notáveis: eBay, Goodwill Industries, National Geographic, Ogio e Telus.

4.2 Plataformas Comerciais

4.2.1 Sisense

É uma ferramenta utilizada por empresas de vários segmentos e atualmente se encontra entre os top 10 da lista de aplicativos de business intelligence, Data Mining e Data Análises. Desde startups e empresas em desenvolvimento até gigantes da Fortune 500, incluindo eBay, Sony, ESPN, Comcast e NASA utilizam Sisense (Sisense, 2017). Esta plataforma ajuda os utilizadores a unificar todos os dados de que precisam e visualizá-los em painéis interativos, transformá-los em 'insights' e compartilhá-los com colegas, parceiros de negócios e clientes através de painéis interativos.

In-Chip uma tecnologia premiada do aplicativo, permite que ele use efetivamente a memória cache da CPU, bem como processe e prepare os dados somente quando uma nova consulta é feita, permitindo que o Sisense escaneie facilmente terabytes de dados e milhares de

utilizadores. Consultas simultâneas e ad-hoc que retornam resultados em segundos, mesmo quando se trabalha com bilhões de linhas de dados em rápida mudança.

O uso de recursos computacionais do mecanismo In-Chip otimiza a necessidade de manter pilhas maciças de hardware ou software. Os dados estão facilmente disponíveis para analistas e administradores, enquanto os profissionais de TI são livres para se concentrar em suas tarefas principais (Sisense, 2017)

O Sisense também permite uma implantação instantânea com a arquitetura Single-Stack do aplicativo que permite desde a integração de dados para a visualização com uma única solução de BI, eliminando a necessidade de usar ferramentas adicionais.

O software ágil de BI cobre o escopo completo de análise de negócio desde a elaboração de dados complexos para análise até a criação de painéis com uma grande variedade de visualizações. Sem a necessidade de um DBA, sem scripts especiais para aprender e sem mistura de manuais para manter.

A API Rest do Sisense pode ser utilizada para integrar facilmente com outros produtos, podendo gerar rapidamente insights de processos de negócios, como gerenciamento de projetos, contabilidade e CRM usando o módulo robusto de análise de dados da Sisense.

A Sisense lançou um recurso inovador ano passado que alavancam o assistente digital pessoal Alexa (PDA) da Amazon para consultas e interpretação habilitadas para voz, bem como bots Sisense e seu roteiro de longo prazo que se concentra na descoberta de dados inteligentes, e na análise de crowdsourcing. Porém no eixo vertical do gráfico de Gartner sua posição é reduzida nas questões de linha de produtos em termos de cloud computing, análises avançadas, menor conscientização do mercado e os riscos competitivos associados a um fornecedor menor em um mercado lotado.

4.2.1.1 Detalhes Técnicos

Tabela 4-Detalhes Técnico Sisense

Dispositivos Suportados	Suporte de Línguas	Modelo de Pagamento	
Windows Android iPhone/iPad MAC Baseado na Web	USA, UK, Canadá, China, India, Japão	Baseado em Cota Inscrição Anual	
Implantação	Tipo de cliente	Satisfação Usuário	Pontuação
Hosp. Nuvem Hosp. Local	Médias e Grandes Empresas	99%	9.7

Cientes Notáveis: NASDAQ, Merck, Dannon, Booking, Comcast, NASA, ESPN, Sony

4.2.2 Tableau

Esta plataforma de fácil usabilidade, com opções de arrastar e soltar e muito intuitiva permite que seus utilizadores criem e publiquem painéis e compartilhe o fluxo de dados de uma plataforma para outra, com seus parceiros ou clientes, sem nenhum obstáculo. Os fornecedores asseguram ter desenvolvido uma solução moderna com um mecanismo de análise de dados 100 vezes mais rápida nas operações do que as de seus concorrentes, permitindo que os utilizadores se conectem a uma série de fontes de dados que já estejam sendo utilizadas. Tableau ganhou o prêmio de melhor Software de Inteligência de Negócios para 2016. (Tableau, 2016).

Tableau divide seus níveis de produto pela implementação e intenção de uso em 3 níveis:

- Tableau Desktop (usado para visualizar e analisar dados, criar pastas de trabalho, visualizações, painéis)
- Tableau Server (usado para editar e distribuir ativos de BI)
- Tableau online (versão hospedada do Tableau Server)

A sua investigação de dados começa com painéis grandes e tradicionais e termina com explorações complexas baseadas em Nuvem. A plataforma permite que o utilizador faça uma combinação de abordagens e visualizações ao invés de realizar análise com um único método, tornando as informações mais ricas e podendo ser facilmente compartilhadas com os clientes e parceiros de diferentes indústrias.

Sua equipe de suporte técnicos de confiança é disponibilizada aos utilizadores e automatiza atualizações de dados para ajudar as empresas a obter novas informações, sem perder o seu tempo (Tableau, 2016) (Tableau, 2017)

O Tableau é um dos três vendedores posicionados no quadrante Líder deste ano. O Tableau continua a ser percebido como o líder moderno do mercado de BI - ainda ligeiramente à frente da Microsoft na execução geral. (Gartner, 2017).

4.2.2.1 Detalhes Técnicos

Tabela 5-Detalhes Técnico Tableau (Tableau, 2017)

Dispositivos Suportados	Suporte de Línguas	Modelo de Pagamento	
Windows Android iPhone/iPad MAc Baseado na Web	USA, Canadá, Internacional	Baseado em Cota Inscrição Anual Pagamento Único	
Implantação	Tipo de cliente	Satisfação Usuário	Pontuação
Hosp. Nuvem Hosp. Local	Médias e Grandes Empresas	93%	9.2

Clientes Notáveis: AeroMexico, SapceX, University of Birmingham, Audi, Banc of America

4.2.3 Microsoft Power BI

Microsoft Power BI é uma ferramenta de BI na Nuvem de autoatendimento fácil e tem uma versão gratuita. O que começou como um suplemento para o ecossistema da Microsoft agora se tornou um pacote totalmente funcional que ocupa o topo do espaço de BI de autoatendimento. Esta ferramenta ajuda os utilizadores a rastrear seus negócios e obter respostas rápidas através de painéis robustos e abrangentes que são disponíveis para todos os dispositivos. Facilita a sua empresa a reunir informações, organizar e conceber estratégias comerciais eficazes através da conversão dos dados em visuais muito atraentes e compreensivos. O sistema é criado para que você identifique as tendências que ocorrem, e para orientar o seu negócio para o sucesso.

O Power BI oferece conectores internos que simplificam a importação de seus dados de soluções como Marketo, Salesforce, Google Analytics e muitos mais e facilita na criação de relatórios simples e publicação rápida.

Microsoft Power BI faz parte do conjunto de produtos empresariais da Microsoft por isso oferece aos utilizadores uma integração com outras ferramentas de negócios populares e amplamente utilizadas, como o Microsoft Excel, Azure e SQL Server. (PowerBI, 2017)

Recentemente foi lançado uma nova versão do Power BI com recursos de relatórios extras que podem ser facilmente incorporados em aplicativos de desenvolvedores personalizados, que podem ser publicados e visualizados facilmente na web com a pretensão de segmentação de mensagens e e-mails, tornando-os acessíveis aos interessados.

Também possui a possibilidade de compra de uma ferramenta Gateway Pessoal usadas para autenticar fontes de dados locais atualmente localizadas fora da firewall. Porém esta ferramenta só pode ser utilizada no S.O Windows que permite a conexão com análise de hospedagem na Nuvem.

4.2.3.1 Detalhes Técnicos

Tabela 6-Detalhes Técnico Power BI

Dispositivos Suportados	Suporte de Línguas	Modelo de Pagamento	
Windows iPhone/iPad Baseado na Web	USA, UK, Canadá, China, Alemanha, Índia, Japão	Vs. Gratuita Pág. Mensal	
Implantação	Tipo de cliente	Satisfação Usuário	Pontuação
Hosp. Nuvem Hosp. Local	Médias e Grandes Empresas	99%	9.1

Cientes notáveis: MediaCom, ABB; Cond Nast, Chevron

5 Avaliação da solução BI

Numa organização coexistem várias aplicações, cada uma com as suas características e integrada à empresa em diferentes níveis de abrangência e profundidade. Um dos desafios que se coloca neste projeto é definir uma estratégia para análise das plataformas proposta no capítulo 3.

Para responder a este desafio neste capítulo propõe-se um método que possibilita uma análise sistemática das várias dimensões sob as quais o problema pode ser visto. Os métodos a ser utilizados consideram-se a avaliação do Instituto de Gartner e sobre este resultado faremos uma análise com critérios qualificativos escolhidos pelos clientes afins de melhor orientar a tomada de decisão na escolha da plataforma BI.

5.1 Avaliação das Plataformas de BI

Sobre a ótica de Mercado de TI pelo grupo Gartner, os fornecedores são julgados pela sua capacidade e sucesso e o resultado desta análise, onde os consumidores acabam comprando essa ideia, passa uma visão diferenciada entregando experiências positivas aos clientes como vendas, suporte, qualidade do produto, habilitação de usuários, disponibilidade, facilidade de atualização / migração e determina a habilidade de execução de um fornecedor. (Gartner, 2017).

A extensa participação do envolvimento de analistas de TI, tanto interno como externo foram tidos em consideração para as avaliações feitas pelos especialistas de Gartner para que as plataformas possam ser manipuladas pelos utilizadores de negócio sem a assistência de TI. Todas plataformas incluídas tiveram várias avaliações, porém o primeiro critério tinha de ser de teor generalizado voltado para o ambiente de BI que tiveram a seguinte classificação: carregar e modelar dados, criar uma camada semântica, construir estruturas de dados como pré-requisito para usar a plataforma de BI ou plataformas de desenvolvimento de TI focadas em construção analítica. (Rita L. Sallam, 2017)

5.1.1 Metodologia Aplicada

A pesquisa feita pelo instituto de Gartner apresenta-se como estudo através de pesquisa de campo, constituindo a pesquisa exploratória por meio de documentos, vídeos, entrevistas e questionários. Esta pesquisa, feita através da visão exclusiva dos pesquisadores, foi interpretar em cinco casos de uso:

- **Agile Centralized BI Provisioning:** Suporta um fluxo de trabalho ágil habilitado para TI, desde dados a conteúdo analítico gerenciado e gerido de forma centralizada, usando os recursos de gerenciamento de dados autônomos da plataforma.
- **Análise descentralizada:** Suporta um fluxo de trabalho de dados para análise de autoatendimento. Inclui análise para unidades de negócios individuais e usuários.
- **Governado Data Discovery:** Suporta um fluxo de trabalho de dados para análises de autoatendimento para SOR, conteúdo gerenciado por TI com governança, reutilização e promoção do conteúdo gerado por usuários para dados SOR certificados e conteúdo analítico.
- **OEM ou Embebido:** Suporta um fluxo de trabalho de dados para conteúdo de BI incorporado em um processo ou aplicativo.
- **Implementação da Extranet:** Suporta um fluxo de trabalho semelhante ao provisionamento ágil e centralizado de Business Intelligence para o cliente externo (Gartner, 2017).

5.1.2 Os 15 critérios de uma plataforma de BI

Para que as classificações das ferramentas fossem possíveis foram destacados 15 critérios fundamentais pelo Instituto de Gartner no âmbito de plataforma de BI na Nuvem. Esta análise pode ser utilizada pelas organizações como sistemas de classificação e medição, para melhor auxiliá-las na sua tomada de decisão. Estes critérios se encontram divididas em 4 subcategorias como: Infraestrutura, Gestão dos dados, Análise e criação de conteúdo e Partilha de resultados.

Infraestrutura

- **Administração, segurança e arquitetura de BI:** todas características que estão relacionadas com segurança da plataforma, administração e auditoria de acesso, garantia de disponibilidade, otimização do desempenho e recuperação de perdas;
- **Cloud BI:** serviço de plataforma e Serviço de aplicativo analítico com capacidades para implementar, implantar e gerenciar análises de dados tanto na Nuvem quanto no local;
- **Conectividade a base de dado:** permite aos usuários uma conexão a dados não estruturados vindos de em diferentes tipos de plataformas de armazenamento, tanto no local como na Nuvem;

Gestão de Dados

- Gerenciamento de meta dados: deve permitir aos administradores uma ferramenta completa e centralizada para que os objetos de metadados possam ser capturados, reutilizados, publicados como dimensões, hierarquias, métricas e indicadores de desempenho, e assim por diante;
- Ferramenta ETL e armazenamento de dados: mecanismo de desempenho autônomo, para trabalhar os dados no acesso, integração, transformação e carregamento de dados e com a capacidade de indexação, gerenciamento de cargas e atualização de agendamento;
- Preparação de dados de Autoatendimento: possibilidade de combinação de dados vindos de diferentes fontes e capacidade de criação de modelos analíticos definidas pelo usuário, como medidas, conjuntos, hierarquias, etc. Nas capacidades avançadas incluem recursos para aprendizagem de máquina, juntas inteligentes, perfil inteligente, geração hierárquica.

Análise e Criação de Conteúdo

- Análise avançada incorporada: permite que o utilizador tenha acesso fácil a recursos avançados de análise que esteja incorporada na ferramenta ou através da importação e integração de modelos externos.
- Painéis ou Dashboards analíticos: a capacidade de criar painéis interativos e conteúdo com exploração visual. E disponibilização para consumos exterior da plataforma de Análise avançadas e aeroespaciais.
- Exploração visual interativa: permite a exploração, visualização e exibição de dados através de vários recursos visuais, como gráficos, barras e linha para incluir mapas de calor, mapas geográficos, gráficos de dispersão, etc.
- Descoberta inteligente de dados: dados relevantes para usuários, como correlações, clusters, links, previsões de dados, são descobertos sem a criação dos modelos ou algum outro tipo de algoritmo. Os dados são explorados pelos utilizadores através de visualizações, narração de linguagem natural, pesquisa e tecnologias NLQ.
- Exploração e criação de dispositivos móveis: capacidade para desenvolvimento e entrega de conteúdos para dispositivos móveis

Partilha de resultados

- Incorporação de conteúdo analítico: possibilidades de incluir componentes de softwares, APIs, suportes para padrões abertos com o objetivo de criar e modificar conteúdos analíticos, visualizações e aplicativos que possam ser incorporados em processo de negócio, aplicativo ou portal. Não tendo a necessidade de residir no aplicativo, mas que devem ser facilmente acessíveis reutilizando a infraestrutura analítica sem obrigar os utilizadores a alternar entre os sistemas.;
- Publicação do conteúdo analítico: funcionalidade que permite aos utilizadores publicar, desenvolver, e operacionalizar conteúdo através diferentes outputs e métodos de distribuição;

Capacidades globais das plataformas

- Capacidades de Plataforma e Fluxo de Trabalho: Este critério analisa as funcionalidades que são fornecidas em um único produto sem emendas ou em vários produtos com pouca integração.
- Facilidade de uso e recurso visual: facilidade de uso para administrar e implantar a plataforma, criar conteúdo, consumir e interagir com o conteúdo, assim como o recurso visual.

No capítulo Anexo tem o resumo em um quadro com os critérios de avaliação e seus pesos juntamente com as forças e fraquezas das plataformas.

5.2 Análise Comparativa das plataformas de BI

Baseado nos resultados obtidos pelo Quadrante, foi possível concluir que para determinados segmentos de organizações, uma ferramenta pode se destacar em relação a outra. Como forma de qualificar as ferramentas, a seleção de Gartner foi dividida em várias fases entre vários critérios com metas a superar e dos quais destacaram-se 15 critérios para a seleção das 24 plataformas, que foram considerados fundamentais para uma ferramenta de BI. De forma a tornar a nossa escolha mais correta possível considerando as necessidades face ao cenário do Departamento informático do Isep.

Vamos fazer uma análise comparativa entre as 5 ferramentas com alguns critérios qualificativos que foram considerados mais importantes pelo cliente. Este método de avaliação simples é citado na norma internacional de avaliação da qualidade de produtos de software, a ISO/IEC 14598-4 (Standardization., s.d.), pela perspectiva do comprador do produto e a ISO/IEC 9126-1 (Standardization., s.d.).

Para avaliarmos o método, cada critério tem uma classificação em função de uma escala qualitativa, compreendida em:

Bom (valor 3): a funcionalidade é indicada como eficiente, excedendo o que era expectável;

Suficiente (valor 2): a funcionalidade se manifesta a um desempenho adequado aos seus propósitos;

Insuficiente (valor 1): a funcionalidade não corresponde ao esperado;

Não será possível garantir através desta avaliação isenção de subjetividade, pois qualquer processo de avaliação possui a sua. No entanto utilizamos este método a fim de delimitar o mais possível a leitura relacionada as funcionalidades mais importantes das ferramentas segundo o cliente.

E os critérios qualitativos foram classificados da seguinte forma:

Escalabilidade: refere-se à capacidade dos utilizadores identificarem, localizarem e executarem as funcionalidades, bem como se apresenta uma interface agradável, intuitivo e ergonómico;

Integração: incorporação com novos módulos, Softwares, Hardwares, Sistemas, etc.

Escalabilidade: aumento de capacidade do sistema, nos dados ou de usuários;

Acessibilidade e Usabilidade: refere-se à capacidade de os utilizadores identificarem, localizarem e executarem as funcionalidades. Também se refere a interfaces mais intuitivas e ergonómicas;

Habilitação ao Utilizador: auxiliam quanto ao treinamento, comunidade, disponibilidade de habilidades no mercado, entre outras.

- Desempenho: refere-se à capacidade de desempenho da ferramenta em seus vários processos;
- Custo: valor do custo da licença, implementação mais ampla do sistema, pacotes, etc.

Tabela 7 - Análise comparativa das Plataformas de BI

Critérios		Plataformas BI				
Peso	Descrição	Pentaho	Domo	Sisense	Tableau	Power BI
3	Escalabilidade	Bom	Bom	Suficiente	Suficiente	Suficiente
3	Integração	Bom	Suficiente	Suficiente	Suficiente	Suficiente
4	Acessibilidade/ Usabilidade	Suficiente	Bom(Nuvem)	Bom	Bom	Bom
2	Habilitação ao Utilizador	Suficiente	Suficiente	Bom	Bom	Bom
4	Desempenho	Suficiente	Bom(Nuvem)	Bom	Suficiente	Bom
5	Custo	Bom	Bom	Suficiente	Suficiente	Bom

Relativamente aos critérios escalabilidade e Integração a plataforma Pentaho apresenta uma classificação Bom, que é considerado acima das expectativas. A Pentaho tem o foco em especial em integração de dados através da PDI, Pentaho Data Integration, para acesso aos dados e recursos analíticos avançados pelo Saiku Analysis. Além do Weka, linguagem R, Implantações avançadas como IoT. Apesar de não serem tratados estes dois últimos itens, e de grande valia destacá-los uma vez que o conceito de Data Mining e IoT são de extrema importância no mercado atual.

O critério usabilidade para uma solução de BI é considerado um item de peso que ajuda muito na escolha de uma ferramenta. É a aderência do produto às necessidades do utilizador, que inclui facilidade de uso em todas as medidas (administração, autores de conteúdo, consumidores de conteúdo e recurso visual). Neste critério temos Domo, Sisense, Tableau e Power BI.

Domo oferece uma experiência de usuário atraente e moderna, alertas extensivos e fluxo de trabalho, com uma variedade de recursos sociais e de colaboração, que se tornam adequadas para a rápida implantação de painéis de controle intuitivos de gerenciamento através da Nuvem nativa. Se destaca entre seus clientes, com facilidade de uso, acesso e integração de dados e rápida implementação.

A Power BI se destaca na facilidade de uso para consumidores de conteúdo e uso composta, o entanto, na facilidade de administração e autoria devem ser melhoradas. Também possui uma alta capacidade de manipular dados de múltiplas fontes com facilidade.

Quanto a Tableau oferece uma experiência de exploração visual altamente intuitiva para usuários de negócios para acessar, preparar e analisar seus dados sem a necessidade de codificação. Isso facilita a experiência de fluxo de trabalho analítico e do poder aos utilizadores para explorar e encontrar informações sobre os dados. Já a Sisense possui facilidade de uso no desenvolvimento de conteúdo Através do ElasticCube, um recurso que suporte várias tabelas de fatos e análises em modelos de dados complexos, usuários mais habilitados e administradores pode facilmente ingerir várias fontes de dados (Hadoop ou um Data Warehouse) com facilidade.

Habilidade do Utilizador, apesar de ter um peso relativamente baixo aqui, é muito importante pois se houver fraqueza neste requisito pode impedir o crescimento da empresa, a satisfação do cliente e o valor da plataforma. Sisense, Tableau e Power BI tem pontuações como Bom. A performance de uma ferramenta tem um peso muito importante, uma vez que um baixo performance pode ser um fator decisivo para a escolha da ferramenta. As ferramentas que saíram a frente foram Domo, Power BI e Sisense.

Os clientes de referência da plataforma Domo informam que o nível de dados que eles podem consultar dentro do repositório de dados do Domo (Domo Cloud) é relativamente alto com ótima performance, porém sua abordagem é toda centrada na Nuvem, que pode não se adequar às organizações com fontes de dados principalmente primárias. A Power BI, Sisense e Tableau se destaca neste requisito na capacidade de manipular dados de múltiplas fontes e dados complexos com bastante desempenho.

Quanto ao critério custo, a Pentaho foi classificada como Bom, e tem sido uma forte opção para empresas que não tem disponibilidade financeira para implementar uma solução de BI. Com sua versão Community que permite que a empresa tenha flexibilidade de iniciar e amplificar seus projetos sem custos de licença que implica risco baixo. Mesmo com investimento na contratação de uma empresa especializada, caso queiram, que também poderá auxiliar com o treinamento da equipe e com o suporte da ferramenta, pode ser uma vantagem pois sua aquisição representa um baixo custo, justamente por ser de código aberto, que significa que a empresa não terá despesas com a aquisição do Pentaho ou de suas licenças. A plataforma Domo apesar de ter sua versão gratuita, no caso de uma implantação mais ampla, cobra uma taxa maior que para a maioria dos outros fornecedores, que deixa esta ferramenta em desvantagem.

E por fim a Power BI que também se classificou como Bom, possui uma versão desktop gratuita, um baixo preço de assinatura por mês de 9,99 ao ano, e o baixo custo total de propriedade (TCO) coloca esta ferramenta em grande vantagem competitiva.

Com base nos resultados das diferentes comparações, observa-se que o Power BI possui uma pontuação bem elevada, tendo em conta que pontuou nos critérios de maiores pesos. A segunda ferramenta que se destaca é a Pentaho com destaque ao Preço. E por fim a Domo que também tem uma pontuação elevada, porém com sua abordagem centrada na Nuvem ficou em desvantagem na escolha.

Tendo em conta as necessidades pretendidas, a ferramenta que melhor se adequa ao problema, é a Pentaho, apesar da Power BI ter uma pontuação mais elevada, se restringe muito aos critérios integração e escalabilidade por ter seus produtos bastante isolados e nenhuma indicação significativa de que isso irá mudar no futuro próximo. Essa plataforma pode se tornar fortemente vinculadas ao Azure e Cortana. Também a incapacidade de implantar cargas de trabalho SSIS em ambientes não-Windows é uma limitação para nossa implementação uma vez que temos uma grande quantidade de SO em Linux. Quanto ao custo relacionado ao Power BI, que foi um de seus grandes destaques na análise comparativa, caso ocorra, as opções adicionais de escala de dados incorrerão em custos adicionais ao alavancar o Microsoft SQL Azure ou o HD Insight na Nuvem - uma vez que atingir os 10GB por limite de usuário no preço padrão do Power BI Pro.

Considerando todas estas desvantagens do Power BI, a ferramenta mais indicada para nosso cenário ficou com a plataforma Open Source Pentaho.

6 Solução Proposta

6.1 Introdução

Antes da implementação de uma solução de BI na Nuvem deve existir uma estratégia de adoção da ferramenta alinhada com a estratégia da empresa. Deste modo tem de se muito bem levantar as informações essenciais sobre o ambiente interno de uma organização como questões vinculadas a área de negócio, arquitetura existente e principalmente ao modelo do serviço a ser aplicado.

O foco desta fase do projeto consiste na análise de negócio de uma solução automatizada dos registos académicos de unidade curricular dos alunos do departamento de Informática do Instituto Politécnico do Porto (ISEP). Compondo todo o cronograma desta fase inclui-se o Escopo do projeto, Analise do Negócio, Análise de Requisitos, Análise de definição do Modelo Dimensional.

Foi utilizado também um segundo caso de uso, o de vendas Motor Inc, para a exploração mais detalhada do processo que engloba BI na Nuvem, numa visão mais técnica entre as várias ferramentas que compõe a Pentaho Suite.

6.2 Âmbito do projeto

Este projeto tem a finalidade na implementação de um modelo dimensional e sua exploração utilizando todo conceito que envolve um Data Warehouse e todos os componentes técnicos de uma implementação de BI. Na fase de análise abordaremos alguns artefactos de engenharia de software e de sistema da informação como o modelo dimensional que avalia a performance do desempenho académico dos alunos do Departamento de Informática do Isep. E na segunda fase a exploração técnica da ferramenta Pentaho e seus componentes. O Sistema deve permitir a exploração de:

- Preparação e processo ETL dos dados;
- Carregamento dos dados em uma área de estagiamento;
- Carregamento dos dados em uma Data Mart (DW);
- Exploração de ferramentas Analíticas, Análises OLAP;

- Exploração e visualização de dados em Relatórios;
- Exploração e visualização de dados em Dashboards com indicadores de desempenhos;
- Compartilhamento de documentos;
- Restrição de dados para perfis de utilizadores (permissão de utilizadores);
- Análise de Resultados;

6.3 Análise do Negócio

Domínio: Solução de uma plataforma de BI na Nuvem, Suite Pentaho, no contexto Educacional.

Departamento de Engenharia Informática, o DEI, hoje com seus 1200 alunos e 78 docentes (23 dos quais doutorados), gera uma quantidade enorme de dados (Moreira, 2015). Neste momento existem algumas plataformas que fazem a gestão da informação para o serviço académicos, mas as informações levantadas dos registos académicos das unidades curriculares dos alunos do Departamento do Isep ainda são tratadas com tecnologias ainda nada dinâmicas. A proposta aqui é a modernização destes serviços, através de ferramentas mais vigorosas, mas ao mesmo tempo fáceis ao nível de manipulação de dados permitindo uma análise multidimensional com visualização dos dados em relatórios e Dashboards aumentando a qualidade e produtividade dos serviços académicos e dando suporte na gestão da informação para a tomada de decisão.

6.4 Definição dos Requisitos

Em geral, o levantamento de requisitos deve compreender o que um software deve e não deve atender. Existem dois tipos de requisitos, os funcionais e os não funcionais. Os requisitos funcionais indicam o que o sistema deve fazer, as suas funções e informações. Os requisitos não funcionais referem aos critérios que qualificam os requisitos funcionais de qualidade para o software, como Usabilidade, performance, confiabilidade, segurança, portabilidade, etc. Em um sistema de Informação os indicadores são considerados também um requisito não funcional, pois sua característica é medir qualidade e desempenho do produto. (Takashina, 1998).

Os requisitos foram levantados através da leitura e Análise da documentação disponibilizada pelo docente Carlos Ferreira, um ficheiro Excel contendo todos os dados a serem explorados que permitiram focalizar e definir as necessidades do negócio. Após a análise dos dados para a identificação dos requisitos, efetuou-se a definição do modelo dimensional para a alimentação do processo ETL na primeira fase do BI.

A nossa base para se chegar aos requisitos funcionais foram analisar as questões antes feitas pelo cliente. Por exemplo: Quais os alunos com as notas mais altas no curso em específico; qual a taxa de alunos reprovados em um determinado curso num dado ano.

Relativamente à especificação dos requisitos, estes são identificados através de um código previamente definido, para melhor estruturação e posterior referências futuras. Foram definidas três categorias de requisitos. Requisitos funcionais, Requisitos Não Funcionais e requisitos de índices de qualidade ou performance ou indicadores.

A codificação segue os seguintes critérios:

Requisitos Não Funcionais: RQNF_00X

Requisitos Funcionais Gerais: RQF_00X

Requisitos Funcionais Indicadores: RQNF_IN00X

Não são classificados aqui os requisitos quanto a sua prioridade, pois considera-se que todos ele tem a mesma prioridade de importância.

6.4.1.1 Requisitos Não Funcionais de software

Tabela 8-Requisitos não funcionais

Identificação	Descrição do Requisito
RQNF_001	A plataforma deverá ser suportada nos sistemas operacionais Windows, Mac, Unix e Linux(32/64Bits);
RQNF_002	A plataforma deverá ser acessada pelos navegadores Internet Explorer (Edge) e Mozilla Firefox ou Google Chrome;
RQNF_003	Os computadores que desejarem acessar a Suite deverão conter a máquina virtual (Java Runtime Environment 1.8 (JRK8 - 32/64Bits) instalada, 8 GBRam (recomendado); 16 GB Espaço em disco(recomendado); Processador32/64Bits Dual-Core ou Core-2-Dual; CPU1.7GHz ou superior;
RQNF_004	Processo ETL, carregamento e atualização da DW e cubo OLAP devem ser automáticos;
RQNF_005	A aplicação a ser desenvolvida deve ser disponibilizada através de uma solução de BI GPL/GNU;
RQNF_006	Todos componentes de BI implementados devem ser disponibilizados na Nuvem;

6.4.1.2 Requisitos Funcionais (Gerais)

Tabela 9-RF Autenticação e Restrição de Acesso

Identificação	RQF_001
Nome	Autenticação e Restrição de Acesso
Fonte Responsável do Requisito	Engenheiro de Requisitos Ana Medeiros
Classificação quanto à Obrigatoriedade	☒ Imprescindível ☐ Desejável
Classificação quanto à prioridade	☒ Alta ☐ Baixa
Especificação do Requisito	A aplicação deve permitir ao utilizador a autenticação através das credenciais (utilizador e password) utilizadas no acesso a quaisquer serviços disponibilizados e restringi-lo segundo seus critérios de acessibilidade.

Tabela 10-RF Termina da Sessão

Identificação	RQF_002
Nome	Termino da Sessão
Fonte Responsável do Requisito	Engenheiro de Requisitos Ana Medeiros
Classificação quanto à Obrigatoriedade	☐ Imprescindível ☒ Desejável
Classificação quanto à prioridade	☐ Alta ☒ Baixa
Especificação do Requisito	O Sistema ao fim de um tempo, por razões de segurança, deverá fazer logout automaticamente.

Tabela 11-RF Navegação Drill Down e Roll Up

Identificação	RQF_003
Nome	Navegação Drill Down e Roll Up
Fonte Responsável do Requisito	Engenheiro de Requisitos Ana Medeiros
Classificação quanto à Obrigatoriedade	☒ Imprescindível ☐ Desejável
Classificação quanto à prioridade	☒ Alta ☐ Baixa
Especificação do Requisito	O Sistema deve permitir o utilizador a fazer consultas OLAP e navegar pelo cubo hierarquicamente ao nível mais baixo de sua granularidade.

Tabela 12-RF Visualização dos dados em Relatórios

Identificação	RQF_004
Nome	Visualização dos dados em Relatórios
Fonte Responsável do Requisito	Engenheiro de Requisitos Ana Medeiros
Classificação quanto à Obrigatoriedade	() Imprescindível (x) Desejável
Classificação quanto à prioridade	() Alta (x) Baixa
Especificação do Requisito	A aplicação deve permitir ao utilizador visualização dos dados em Relatórios.

Tabela 13-RF Visualização dos dados em Dashboards

Identificação	RQF_005
Nome	Visualização dos dados em Dashboards
Fonte Responsável do Requisito	Engenheiro de Requisitos Ana Medeiros
Classificação quanto à Obrigatoriedade	(x) Imprescindível () Desejável
Classificação quanto à prioridade	(x) Alta () Baixa
Especificação do Requisito	A aplicação deve permitir ao utilizador a visualização dos dados trabalhados em Dashboards.

Tabela 14-RF Visualização dos dados em Gráficos e Tabelas

Identificação	RQF_006
Nome	Visualização dos dados em Gráficos e Tabelas
Fonte Responsável do Requisito	Engenheiros de Requisito Ana Medeiros
Classificação quanto à Obrigatoriedade	(x) Imprescindível () Desejável
Classificação quanto à prioridade	(x) Alta () Baixa
Especificação do Requisito	A aplicação deve permitir ao utilizador visualizar os dados apresentados em tabelas e Gráficos.

Tabela 15-RF Agendamento de Execução de Tarefas

Identificação	RQF_007
Nome	Agendamento de Execução de Tarefas
Fonte Responsável do Requisito	Engenheiro de Requisitos Ana Medeiros
Classificação quanto à Obrigatoriedade	() Imprescindível (x) Desejável
Classificação quanto à prioridade	() Alta (x) Baixa
Especificação do Requisito	A aplicação deve permitir ao utilizador a agendar tarefas a serem executadas em horários ou dias específicos.

Tabela 16-RF Análise Olap

Identificação	RQF_008
Nome	Análise Olap
Fonte Responsável do Requisito	Engenheiro de Requisitos Ana Medeiros
Classificação quanto à Obrigatoriedade	() Imprescindível (x) Desejável
Classificação quanto à prioridade	() Alta (x) Baixa
Especificação do Requisito	A aplicação deve permitir ao utilizador a fazer Análise de dados em cubos Olap.

Tabela 17-RF Exportar Dados para tipos de ficheiros diferentes

Identificação	RQF_009
Nome	Exportar Dados para tipos de ficheiros diferentes
Fonte Responsável do Requisito	Engenheiro de Requisitos Ana Medeiros
Classificação quanto à Obrigatoriedade	() Imprescindível (x) Desejável
Classificação quanto à prioridade	() Alta (x) Baixa
Especificação do Requisito	A aplicação deve permitir ao utilizador exportar as informações da camada de visão para ficheiros de formato Excel, PDF, World ou CSV.

6.4.1.3 Requisitos Não Funcionais ou Indicadores de Performance

Estes requisitos foram retirados depois de ter feito a análise do ficheiro Excel e são considerados muito importantes para a medição d

Tabela 18Tabela de Indicadores de Performance

Identificação	Definição
RQNF_IN01	Taxas de aprovação dos Aluno por Disciplina
RQNF_IN02	Média das notas por Disciplina
RQNF_IN03	Media total notas Disciplinas por semestre/ano
RQNF_IN04	Media das notas dos Alunos por semestre/ano
RQNF_IN05	Quantidade total/Percentagem das notas de UCs em que o aluno obteve aprovação por ano.
RQNF_IN06	Quantidade total/Percentagem de alunos que não obtiveram aprovação por disciplina.
RQNF_IN07	Percentagem de UCs a que cada aluno esteve inscrito e que obteve aprovação.
RQNF_IN08	Percentagem de UCs a que cada aluno esteve inscrito e que obteve reprovação.
RQNF_IN09	Percentagem de UCs em que o aluno obteve aprovação por ano e por horário (laboral/pós-laboral desc.).
RQNF_IN10	Percentagem de unidades curriculares com aprovação dos alunos do 1º ano
RQNF_IN11	Percentagem de UCs com aprovação por período/semestre.
RQNF_IN12	Percentagem de alunos com 0 aprovações por semestre/ano
RQNF_IN13	Total/Percentagem de alunos que faltam para concluir o curso.
RQNF_IN14	Total de Alunos inscritos.
RQNF_IN15	Resultados dos alunos com nota de frequência positiva.

6.4.2 Utilizadores e stakeholders

Diretores, subdiretores e professores.

6.5 Definição do Modelo Dimensional

De acordo com o modelo multidimensional, que tem como formação elementos básicos como tabela fato, dimensões e medidas é uma técnica de conceção e visualização de um modelo de dados que descrevem aspetos de negócios. É utilizado para a sumarização e estruturação dos dados para a apresentação dos mesmos afim de que se aplique análises de seus valores. (MACHADO, 2007).

6.5.1 Metodologia Kimball para modelação de um DW

Kimball contrário ao modelo relacional, embora enfatize essencial a variabilidade deste modelo, defende o modelo dimensional para a análise de dados, e a sua proposta metodológica, no entanto, se baseada inteiramente no modelo relacional, que se fez a prescrição de regras de implementação do modelo gerenciador de um Sistema de BD. A proposta de Kimball é representado por um Schema em estrela onde existe uma tabela central, a tabela de facto, que contém os factos ocorridos e possui ligações com as dimensões através de suas chaves. A dimensão por sua vez contém características de um evento. (Gomes, 2013).

O modelo dimensional escrito por Kimball, foi escrito como uma defesa da legitimidade da utilização de dados desnormalizados, visando a prática que assegura o funcionamento diário do negócio e sistema de Analise OLAP, onde o objetivo principal é retirar informações estratégicas. Como tal, a implementação de um design DW tornou-se consensual na modelação de dados próprios, que consiste na conceção intuitiva para utilizador de que a individualidade de qualquer negócio pode ser sempre generalizada tendo em conta o modelo: Existem várias dimensões com vários níveis de hierarquias, e entre suas junções ocorrem factos mensuráveis ou calculáveis, podendo serem analisados seus valores. (Kimball & Ross, 2002).

.

6.5.2 Data Mart na área Gestão Académica

6.5.2.1 Processo do Negócio

Gestão Acadêmica no processo de avaliação do desempenho das notas dos Alunos.

O foco do problema em análise centra-se no acompanhamento dos serviços de Gestão Acadêmica no processo de avaliação dos dados históricos curriculares (das notas/Disciplinas) dos Alunos do Departamento de Informática do Isep entre o ano de 2000 e 2013.

O ficheiro pode ser visto em parte na figura 53 em Anexo.

6.5.2.2 Nível de Granularidade

Vamos considerar para o nível mais baixo de granularidade a verificação das notas dos alunos por semestre, pois cada disciplina é concluída por semestre.

6.5.2.3 Identificação das Dimensões e tabela de Facto Envolvidas

De acordo com a disponibilização dos dados no ficheiro, tendo em consideração que muitos atributos já haviam sido calculados, devemos fazer algumas observações relevantes dos atributos do ficheiro que foi disponibilizado:

- AnoAluno (coluna 14,) e AuxAno (coluna 17), possuem os mesmos valores (1-3) e se referem a mesma informação que é o ano em que o aluno se encontrava fazendo aquela disciplina;
- NDAux (coluna 20):Quantidade de vezes em que o aluno fez as disciplinas (conta as vezes repetidas);
- NDisc (coluna 21): O total das disciplinas concluídas;
- NDApAux (Coluna 22): Possui um contador que começa com 0 e avança de acordo em que as disciplinas vão sendo aprovadas, se a disciplina é reprovada, o contador se repete;
- PDAp (coluna 24): percentagem das disciplinas aprovadas. representa uma medida e não precisa de ser calculada na tabela de facto
- RES (coluna 19): Esta coluna representa todos os status de aprovação, sendo que a Coluna 25 que tem como valores Aprovado e Reprovado, está contido nesta coluna, e as colunas sobre notas também que incluem os valores SMR, SMS, NC, NF, FT, SMNF.
- As colunas NrFreq (coluna 3), Exame (coluna 4) e Nota (coluna 5), possuem campos do tipo numérico e String, sucessivamente. Isso pode ser resolvido com o processo de ETL, ao carregar a dimensão Disciplina.

Definição das Dimensões e Tabela de Facto

DIM_ALUNO: Esta dimensão detém os detalhes sobre o número e nome do Aluno. Ano de entrada do Aluno na instituição e o ano correspondente ao Curso (de 1 a 3).

Esta foi a opção escolhida pois não existem dados suficiente para criar uma dimensão Curso, e como existe dois Anos distintos (Ano Entrada e Ano do Curso),e estão diretamente relacionados ao aluno, então foi tomada esta decisão de inserir estes dois atributos nesta dimensão.

Atributos: PK_DIM_ALUNO; NUMERO-coluna1; NOME-coluna2; ANO_ENTRADA_ALUNO-coluna11; ANO_ALUNO_CURSO-coluna14;

DIM_NOTA: Esta dimensão representa todas as notas dos alunos que se encontra no ficheiro. Ela também poderia ser representada na tabela de facto como Dimensão Degenerada (DD) representada somente pela coluna nota final do qual trás o cálculo de todas as notas.

Atributos: PK_DIM_NOTA; NOTA_FREQ-coluna3; NOTA_EXAME-coluna4; NOTA_FINAL-coluna5; NOTA_FREQ_ARRED-coluna26.

DIM_REGIME: Esta dimensão representa os regimes Parcial, Integral, Extra-Curricular e Gratuito em que o aluno foi inscrito.

Atributos: PK_DIM_REGIME, REGIME-coluna8;

DIM_TURMA: Detém informações sobre a Turma como nome e o turno da Turma.

Atributos: PK_DIM_TURMA, NOME_TURMA; TURNO_TURMA;

DIM_DISCIPLINA: Esta dimensão contém todos os dados relacionados a disciplinas como O código da disciplina, tipo de aprovação, anos avançados, horário, ano e semestre.

Nesta dimensão temos valores como total de disciplinas aprovadas e atrasadas, este tipo de informação poderia ser inserido na tabela agregada, porém nos optamos por deixá-la na tabela de disciplina pois não se trata de uma grande quantidade de dados a serem tratados que exigiria uma solução mais performática. Poderia colocar o curso na disciplina uma vez que o meu curso é Licenciatura ou Mestrado que esta ligado diretamente ao Horário que ser de dia ou de noite

Atributos: PK_DIM_DISCIPLINA-col6; CODIGO_DISCIPLINA-coluna6; TIPO_APROVACAO-col19; DISCIPLINAS_ANOS_AVANCADOS-coluna18; HORARIO-col9, ANO_DISCIPLINA-coluna12, SEMESTRE_DISCIPLINA-coluna13,

FACTO_NOTA_AVALIACAO: A tabela de facto contém as chaves estrangeiras de todas as dimensões, o ano em que o aluno se encontra no curso. Esta coluna é conhecida como Dimensão Degenerada pois não possui uma Dimensão própria. A facto possui ainda as médias das notas e o total das disciplinas.

Atributos: FK_DIM_ALUNO; FK_DIM_DISCIPLINA; FK_DIM_REGIME; FK_DIM_TURMA; PK_DIM_NOTA; ANO_ALUNO_CURSO(DD); MEDIA_NOTAS_EXAME_DISCIPLINAS; MEDIA_NOTAS_FREQ_DISCIPLINAS; MEDIA_APROVACAO_ALUNO; MEDIA_NOTAS_FINAIS_DISCIPLINAS; TOTAL_DISCIPLINAS_APROVADAS-coluna22; TOTAL_DISCIPLINAS_ATRASADAS-coluna16; TOTAL_DISCIPLINA_INACABADAS; TOTAL_DISCIPLINAS_REPROVADAS; NUMERO_DISCIPLINAS_INSCRITAS-coluna21; PERCENT_DISCIPLINAS_APROVADAS-coluna 24;

6.5.2.4 Identificação dos Factos Envolvidas

Os factos ou medidas a serem calculadas podem ser as mesmas indicadas na tabela indicadores de performance. Alguns deles descritos abaixo.

- A média de frequência das Notas dos Exames por Disciplina;
- Percentagem de Aprovações em Disciplinas obtidos pelo aluno por semestre/ano;
- Número de disciplinas em que o aluno foi inscrito e não terminou;
- Número de Disciplinas em que o aluno foi aprovado;
- Número de Disciplinas em que o aluno foi reprovado;
- Percentagem de aluno aprovação por Disciplina por ano.

6.5.2.5 Modelo Dimensional

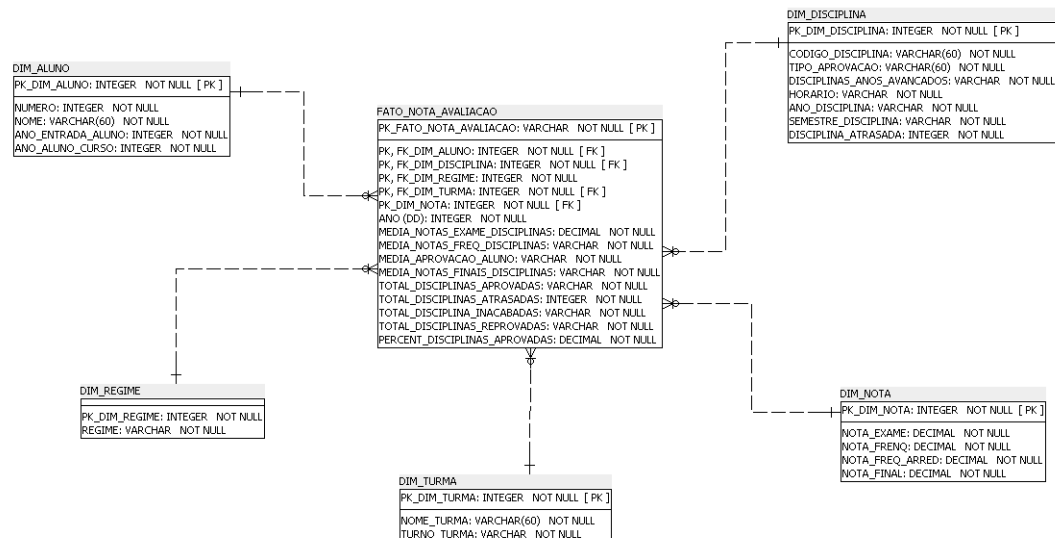


Figura 4- Modelo dimensional gestão do Histórico de notas dos alunos

6.5.3 Data Mart na área Gestão de Vendas.

Para a construção do DW para este canal existem problemas como a falta de informação para gerar estatísticas das lojas que mais vendem, dos funcionários mais esforçados, dos produtos mais vendidos e até mesmo do período que mais tem movimento. Estes dados foram disponibilizados pelo professor Ricardo Gouveia da Empresa Opening, que me orientou na implementação deste modelo.

Na situação hipotética, teremos que construir nosso Data Mart da seguinte maneira:

6.5.3.1 Processo de Negócio

Indústria: Automobilística

Empresa: Motor Inc.

Canal de Vendas : Vendas indiretas

6.5.3.2 Nível de Granularidade

Qual o número total de carros vendidos ao mês por representante e por estado.

6.5.3.3 Definição das Dimensões

Dimensão Produto; Dimensão Vendedor; Dimensão Cliente; Dimensão Tempo.

6.5.3.4 Definição da tabela de Facto

Receita de Vendas e quantidade de vendas.

6.5.3.5 Indicadores de performance

Vamos destacar abaixo alguns indicadores de performance para este cenário:

- Quais foram os 10 modelos mais vendidos este mês?
- Como foram as vendas destes 10 modelos comparados com os 10 mais vendidos nos últimos 6 meses
- Mostre-me as Vendas por representante, modelo e região;
- Qual o número total de carros vendidos dia a dia por representante e estado

Estes indicadores de performance definidos aqui são meramente para efeito de ilustração, não serão trabalhados neste projeto. Os outros índices trabalhados estão definidos no capítulo do Saiku Analytics pois o intuito foi mostrar mais a nível de teste o comportamento da ferramenta diante da análise Olap.

6.5.3.6 Modelo Relacional

Este modelo representa a Base de Dados de produção do qual os dados de origem da empresa Motor Inc fazem parte e que foram disponibilizados para a montagem deste modelo.

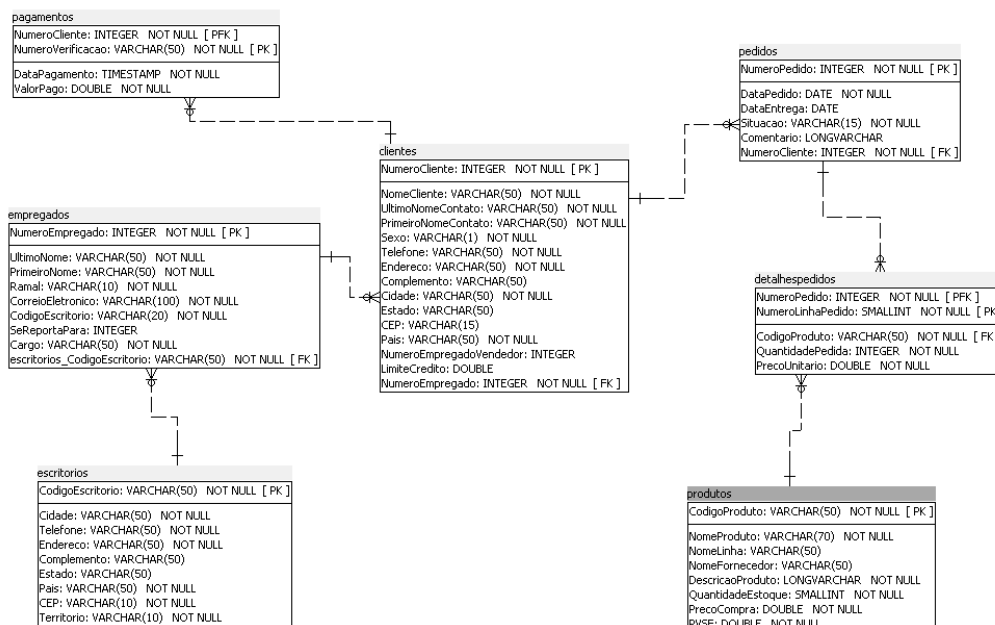


Figura 5-Modelo Relacional Motor Inc.

6.5.3.7 Modelo Dimensional

Este é o modelo dimensional resultante do modelo relacional disponibilizado. Este modelo possui uma estrutura simples da desnormalização dos dados para uma melhor compreensão.

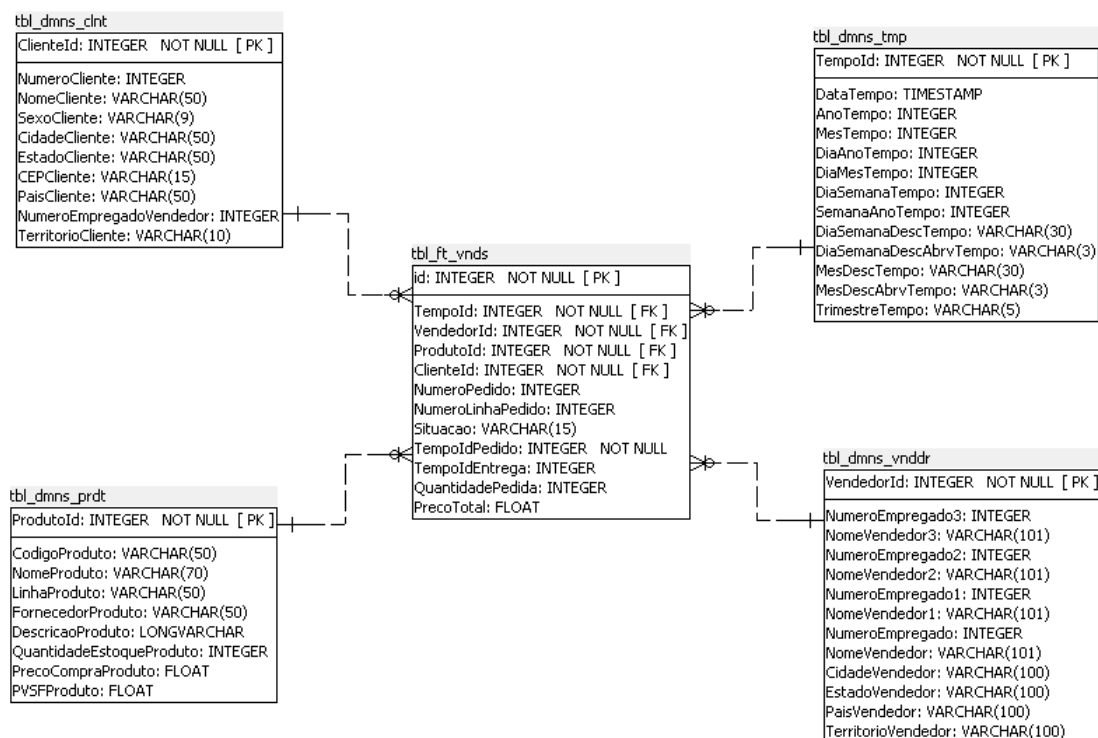


Figura 6-Modelo Dimensional Motor Inc.

6.6 Visão da Implantação – Arquitetura Pentaho BI

Segue abaixo a apresentação do modelo arquitetural implementado que compõe as ferramentas de integração da solução Pentaho utilizadas neste projeto, apresentando-se assim os principais componentes do Pentaho BI na Nuvem.

Esta é uma arquitetura que foi utilizada neste projeto, contendo os seguintes componentes. A BD de produção que pode ser conectada por todos os componentes da Suite e pelo próprio Pentaho BI Server através do JNDI ou SGBD. O Pentaho Data Integration (PDI) com seu processo ETL. A ferramenta de análise cubos OLAP Mondrian (PSW) que utiliza o cliente Saiku para a publicação e construção dos cubos Olap no PUC. O Report Designer (PRD), um programa de desktop que é utilizado para construir o modelo de relatório e que permite emitir solicitações para servidor (Pentaho) para publicação dos relatórios. Ainda no BI Server o CDE, junto com CDA e CDF para a execução dos Dashboards. E a integração do ficheiro Excel com o PDI para que fosse possível a extração dos dados.

A idealização desta arquitetura partiu da necessidade da implementação dos principais componentes que integram a suite Pentaho. Nas seguintes fases das implementações do processo de BI serão mostradas as tecnologias envolvidas e como foram implementadas, consolidando compreensão deste modelo.

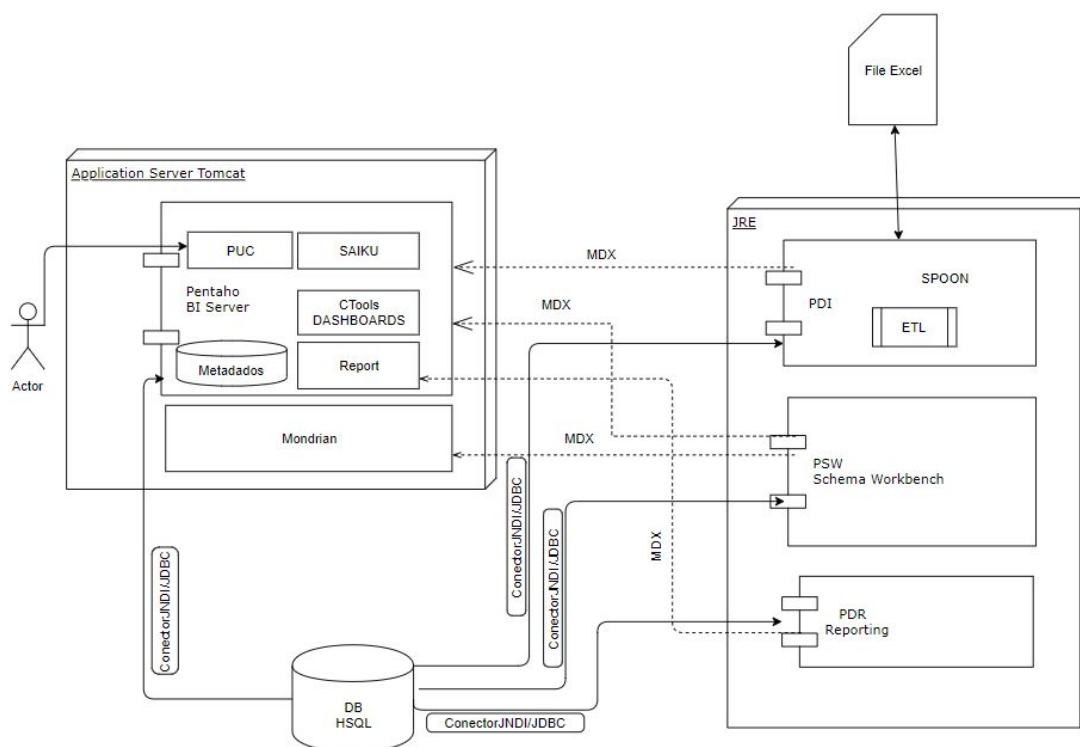


Figura 7-Arquitetura Geral da Implantação Pentaho BI

6.6.1 Definição das tecnologias de integração com Pentaho BI

PSW- Schema Workbench: Ferramenta visual para projetar e testar esquemas de cubo Mondrian. Mondrian usa esses esquemas de cubo para interpretar o MDX e traduzi-lo em consultas SQL para recuperar os dados de um RDBMS.

Saiku é um mecanismo OLAP puro e relacional que trabalha com seus dados armazenados em vez de fornecer seu próprio modelo de armazenamento de dados multidimensional.

PRD - Pentaho Report Designer: O Pentaho Report Designer é o front-end gráfico para criar, editar e publicar relatórios para a plataforma Pentaho BI. Uma das principais vantagens de usar modelos de metadados Pentaho como fontes de dados. Os relatórios também podem ser publicados diretamente no servidor Pentaho para o uso do Pentaho User Portal. (Bouman & Dongen, 2009).

PDI- Pentaho Data Integration: Este mecanismo de integração de dados pode-se considerar como o coração do Pentaho. É capaz de interpretar e executar Jobs e transformações,

realizando assim tarefas reais de integração de dados. (Dongen, 2009). A principal diferença entre jobs e transformations é que os jobs consistem em entradas de trabalho, que representam a execução de uma tarefa, enquanto as transformações consistem em etapas de transformação que representam uma operação em fluxos de registros.

Spoon é a interface gráfica que se encontra no PDI para a montagem das transformações e Jobs.

Um detalhe importante na interface Spoon que deve ser destacado é a execução do fluxo com sucesso. Ele mostra a informação através de check-ins verdes em cada step que compõe o fluxo. Caso o fluxo tenha algum step que não são executados com sucesso, ele interrompe o processo. Subsequentemente pode-se ver nas abas da interface as Métricas, o Histórico da Execução, os logs, a performance dos gráficos e a visualização dos dados. O backtrace de todo o processo que pode ser verificado para os diversos resultados.

PUC: Pentaho vem com uma interface web embutida chamada console do usuário. O console do utilizador forma um front-end que permite que o usuário interaja com o servidor.

CTool

- CDF: permite que os usuários criem, de maneira rápida e fácil, painéis dinâmicos que permitem aos usuários explorar e entender grandes quantidades de dados usando uma variedade de gráficos, tabelas e outros componentes e, em seguida, "detalhar" os dados exatos;
- CDE: é um editor de painel gráfico que fornece acesso aos componentes do painel no CDF.
- CDA: permite coletar e combinar dados de várias fontes de dados em uma única estrutura, que pode ser usada em painéis
-

6.7 Visão da Implementação com Pentaho

A solução será testada em um conjunto de dados de origem de uma Base de dados relacional, que representa as atividades de vendas de uma empresa denominada Motor-Inc., procurando responder a um cenário mais complexo com a exploração das funcionalidades da solução na área comercial.

6.7.1 Implementação do Data Mart de Vendas no processo ETL

O data Mart de Vendas da empresa Motor Inc faz vendas de produtos automobilísticos e atual em vários países do mundo.

Existem algumas inconsistências dos dados na BD Motor Inc que devem ser trabalhados, como a falta de informação sobre onde se localizam as vendas por vendedor. Alguns dados devem sofrer o processo de limpeza, normalização e outros processo do ETL. Também existem uma tabela com autorelacionamento que deve ser aplicado algumas técnica de tratamento de dados

como o SCD e o Role playng Dimension, neste caso para a dimensão tempo. Ao logo das implementações iremos explicando em detalhes os processos aplicados.

O processo que se apresenta a seguir aborda basicamente três fases, a extração dos dados de origem na BD relacional e o carregamento das tabelas temporárias utilizando o conceito de Área de estagiamento. O tratamento e a limpeza dos dados que contribuiu para eliminar inconsistências da Base de Dados como por exemplo, completar dados que faltavam, tratar valores nulos, e eliminar registros irrelevantes para análise. E por fim o carregamento das dimensões no Data Warehouse.

Os dados de origem aqui trabalhados foram disponibilizados em dois scripts para que fosse gerado as tabelas na DB motr-inc-oltp, que será a nossa fonte de dados de produção. Seguidamente foi criada a BD motor-inc-star, onde se encontra duas tabelas temporárias, quatro dimensões e uma tabela de facto. As tabelas tiveram que ser criadas antecipadamente, com nomes semelhantes aos nomes das tabelas de produção. Este procedimento foi seguido para que no momento da execução do processo as tabelas possam ser identificadas para o match entre elas.

O step escolhido para cada transformação vem de encontro com os critérios do negócio, porém poderia ter sido escolhido qualquer outro que atendesse aos requisitos especificados, para obtenção dos mesmos resultados.

Para a realização do carregamento de cada tabela na Base de dados terá que ser processado uma Transformação da interface gráfica Spoon do PDI.

6.7.1.1 Processo de Carregamento da tabela stage_territorio no SA

As tabelas stage_territorio e stage_pais_territorio foram criadas para obter dados geográficos dos clientes, uma vez que os clientes não possuem estes dados guardados em sua tabela.

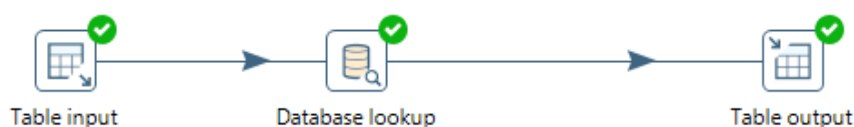


Figura 8-Transformação-Carregamento da dimensão Território na Staging Area

O modelo do negócio motor inc exige informações de vendas por territórios de seus clientes. Logo o objetivo desta transformação é juntar as informações de vendedor que estão vinculadas a um escritório que por sua vez está vinculado à um território. Estes dados serão extraídos da tabela empregado, serão processados e carregados na tabela temporária stage_territorio. Para

o início da transformação o step Table input busca os dados sobre o número do empregado e o código do escritório na tabela empregado através de uma Query. O Database lookUp é responsável para trazer os valores do território ao comparar os códigos de escritório entre as tabelas escritório e empregado. Por fim o Table output carrega os dados na tabela stage_territorio com os valores do NumeroEmpregado, CodigoEscritorio e Territorio.

Logo abaixo podemos ver a Query da Table Input e o resultado do carregamento da tabela stage_territorio na BD.

```
SELECT
  NumeroEmpregado
, CodigoEscritorio
FROM empregados
```

Na Figura 10 podemos ver o resultado do carregamento das tabelas stage_territorio.

6.7.1.2 Processo de Carregamento da tabela stage_pais_territorio no SA

Para o complemento da informação no território, esta transformação busca da informação do cliente, como País e complementa com informações vindas da tabela vendedor, como o território. O resultado desta transformação será utilizado para compor com os outros dados de clientes no processo de carga da dimensão cliente.

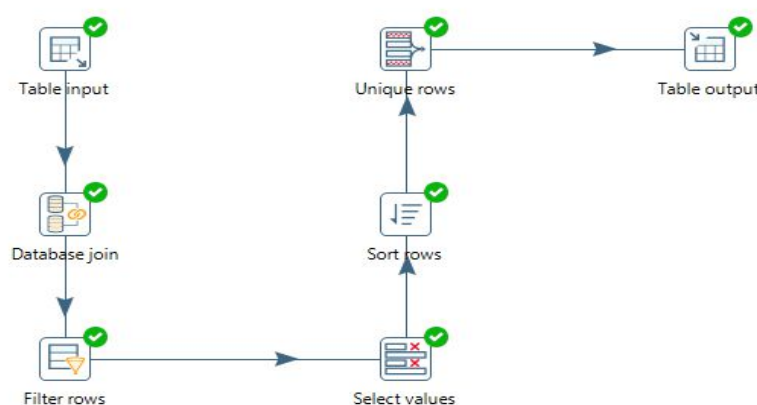


Figura 9-Transformação-Carregamento da dimensão Cliente Território na Staging Area

O step Table input seleciona o atributo País e NumeroEmpregadoVendedor da tabela cliente. O Database join faz join entre as tabelas de origem clientes e a tabela stage_territorio com o objetivo de filtrar o NumeroEmpregado igual ao NumeroEmpregadovendedor. O Filter Row filtra os valores não nulos das tabelas, funciona como um outer join nas queries. O Select Value é utilizado para remover o campo NumeroempregadoVendedor pois este atributo só serviu para fazer a passagem de parâmetro no Database join e também para demonstrar mais uma

funcionalidade deste step. Estes dois últimos procedimentos podem ser aplicados no conceito de data cleaning.

Unique Rows faz uma normalização dos dados ao tirar a duplicação dos dados, porém este step precisa que os dados estejam classificados antecipadamente em ordem alfabética, neste caso o Sort Rows trata disto, e faz a classificação dos atributos de acordo com os critérios, País e território e para finalizar o Table output grava os dados na tabela stage_pais_territorio.

Query da Table input

```
SELECT
País,
NumeroEmpregadoVendedor
FROM clientes
```

NumeroEmpregado	CodigoEscritorio	Territorio
1.002	1	NA
1.056	1	NA
1.076	1	NA
1.088	6	APAC
1.102	4	EMEA
1.143	1	NA
1.165	3	LATAM
1.166	1	NA
1.188	2	NA
1.216	2	NA
1.286	2	NA
1.323	1	NA
1.337	4	EMEA
1.370	4	EMEA
1.401	4	EMEA
1.501	7	EMEA
1.504	7	EMEA
1.611	6	APAC
1.612	6	APAC
1.619	6	APAC
1.621	5	Japão
1.625	5	Japão
1.702	4	EMEA

País	Territorio
Alemanha	EMEA
Austrália	APAC
Brasil	NA
Bélgica	EMEA
Canadá	NA
Dinamarca	EMEA
Espanha	EMEA
Estados Unidos	LATAM
Estados Unidos	NA
EUA	LATAM
EUA	NA
Filipinas	Japão
Finlândia	EMEA
França	EMEA
Hong Kong	APAC
Irlanda	EMEA
Itália	EMEA
Japão	Japão
Noroega	EMEA
Reino Unido	EMEA
Singapura	APAC
Suécia	EMEA
Suíça	EMEA
Áustria	EMEA

Figura 10- Tabela stage_territorio e stage_pais_territorio na BD motor-in-star

6.7.1.3 Processo de carregamento dimensão Cliente

Vamos apoiar nas tabelas da área de estagiamento para complementar a dimensão cliente.

Nesta transformação iremos inserir informações de território e país na tabela cliente para que as informações sobre localidade do cliente possam ser também consultadas.

O Table Input busca as informações vindas da tabela cliente da BD de origem, consecutivamente faz-se a utilização de dois Value Mapper para fazer a limpeza dos dados. O primeiro Value Mapper substitui os valores F e M do atributo Sexo para Feminino e Masculino, respectivamente, e o segundo substitui Estados Unidos para EUA do atributo País. O Database lookUp faz um lookup do País entre as tabelas clientes e stage_pais_territorio que terá como retorno da consulta os territórios.

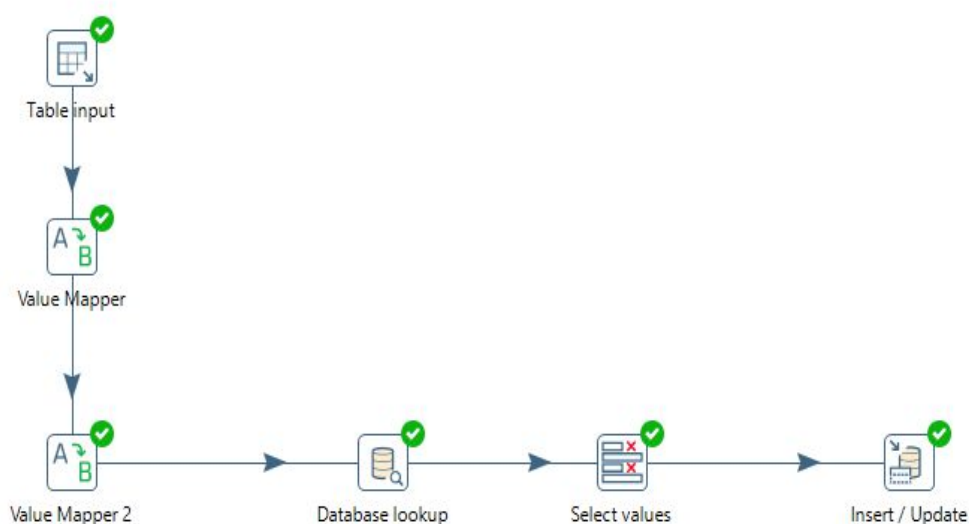


Figura 11-Transformação- Carregamento da dimensão Cliente no Staging Area

No step Select values vamos remover toda informação relacionada diretamente ao cliente (UltimoNomeContacto, PrimeiroNomeContacto, Telefone, Endereco, Complemento, LimiteCredito). Estes atributos pessoais não serão necessários para a análises. Finalizando, o Insert/Update passa os valores no lookup NumeroCliente da dimensão cliente e a tabela cliente de origem. inserindo os dados caso não existam, caso existam ele faz Update nos valores dos campos passado em parâmetros, somente na condição se aquele dado for diferente, ou seja, existe uma mudança a ser feita. Aqui aplicasse o conceito de Slow change Dimension tipo 1. E carrega os valores na dimensão cliente.

Query da Table input.

```

SELECT
  NumeroCliente
, NomeCliente
, Upper (UltimoNomeContato) UltimoNomeContato
, Upper (PrimeiroNomeContato) PrimeiroNomeContato
, Telefone
, Endereco
, Complemento
, Upper (Cidade) Cidade
, Sexo
, Upper (Estado) Estado
, CEP
, Upper (Pais) Pais
, NumeroEmpregadoVendedor
, LimiteCredito
FROM clientes

```

motor-inc-star.tbl_dmns_clnt: 122 registros totais (aproximadamente)

ClienteId	NumeroCliente	NomeCliente	SexoCliente	CidadeCliente	EstadoCliente	CEPCliente	PaisCliente	NumeroEmpregadoVendedor	TerritorioCliente
1	103	Atelier graphique	FEMININO	NANTES	(NULL)	44000	FRANÇA	1.370	EMEA
2	112	Signal Gift Stores	FEMININO	LAS VEGAS	NV	83030	EUA	1.166	LATAM
3	114	Australian Collectors, Co.	MASCULINO	MELBOURNE	VICTORIA	3004	AUSTRÁLIA	1.611	APAC
4	119	La Rochelle Gifts	FEMININO	NANTES	(NULL)	44000	FRANÇA	1.370	EMEA
5	121	Baane Mini Imports	MASCULINO	STAVERN	(NULL)	4110	NORUEGA	1.504	EMEA
6	124	Mini Gifts Distributors Ltd.	FEMININO	SAN RAFAEL	CA	97562	EUA	1.165	LATAM
7	125	Havel & Zbyszczek Co	MASCULINO	WARSZAWA	(NULL)	01-012	POLÓNIA	(NULL)	(NULL)
8	128	Blauer See Auto, Co.	MASCULINO	FRANKFURT	(NULL)	60528	ALEMANHA	1.504	EMEA
9	129	Mini Wheels Co.	FEMININO	SAN FRANCISCO	CA	94217	EUA	1.165	LATAM
10	131	Land of Toys Inc.	FEMININO	NYC	NY	10022	EUA	1.323	LATAM
11	141	Euro+ Shopping Channel	MASCULINO	MADRID	(NULL)	28034	ESPAÑA	1.370	EMEA
12	144	Volvo Model Replicas, Co	FEMININO	LULEÅ	(NULL)	S-958 22	SUÉCIA	1.504	EMEA
13	145	Danish Wholesale Imports	MASCULINO	KOBENHAVN	(NULL)	1734	DINAMARCA	1.401	EMEA
14	146	Saveley & Henriot, Co.	FEMININO	LYON	(NULL)	69004	FRANÇA	1.337	EMEA
15	148	Dragon Souvenirs, Ltd.	MASCULINO	SINGAPORE	(NULL)	079903	SINGAPURA	1.612	APAC
16	151	Muscle Machine Inc	MASCULINO	NYC	NY	10022	EUA	1.286	LATAM
17	157	Diecast Classics Inc.	MASCULINO	ALLEN TOWN	PA	70267	EUA	1.216	LATAM
18	161	Technics Stores Inc.	MASCULINO	BURLINGAME	CA	94217	EUA	1.165	LATAM
19	166	Handy Gifts & Co	MASCULINO	SINGAPORE	(NULL)	069045	SINGAPURA	1.612	APAC
20	167	Hersku Gifts	MASCULINO	BERGEN	(NULL)	N 5804	NORUEGA	1.504	EMEA
21	168	American Souvenirs Inc	FEMININO	NEW HAVEN	CT	97823	EUA	1.286	LATAM
22	169	Porto Imports Co.	FEMININO	LISBOA	(NULL)	1756	PORTUGAL	(NULL)	(NULL)
23	171	Daedalus Designs Imports	MASCULINO	LILLE	(NULL)	59000	FRANÇA	1.370	EMEA
24	172	La Corne D'abondance, Co.	FEMININO	PARIS	(NULL)	75012	FRANÇA	1.337	EMEA
25	173	Cambridge Collectables Co.	MASCULINO	CAMBRIDGE	MA	51247	EUA	1.188	LATAM
26	175	Gift Depot Inc.	FEMININO	BRIDGEWATER	CT	97562	EUA	1.323	LATAM
27	177	Osaka Souvenirs Co.	MASCULINO	OSAKA	OSAKA	530-0003	JAPÃO	1.621	Japão
28	181	Vitachrome Inc.	MASCULINO	NYC	NY	10022	EUA	1.286	LATAM
29	186	Toys of Finland, Co.	MASCULINO	HELSINKI	(NULL)	21240	FINLÂNDIA	1.501	EMEA
30	187	AV Stores, Co.	FEMININO	MANCHESTER	(NULL)	EC2 5NT	REINO UNIDO	1.501	EMEA
31	189	Clover Collections, Co.	MASCULINO	DUBLIN	(NULL)	2	IRLÂNDIA	1.504	EMEA
32	198	Auto-Moto Classics Inc.	FEMININO	BRIKHAVEN	MA	58339	EUA	1.216	LATAM
33	201	UK Collectables, Ltd.	FEMININO	LIVERPOOL	(NULL)	WX1 6LT	REINO UNIDO	1.501	EMEA

Figura 12 Tabela Cliente na BD motor-in-star

6.7.1.4 Processo de carregamento dimensão Tempo

Objetivo desta transformação é apresentar a construção de um processo que irá popular a dimensão tempo agrupando os dados por ano, mês, dia do ano, dia do mês, dia da semana, semana do ano e trimestre. Possibilitando assim as consultas de vendas neste nível de granularidade. Foram inseridas informações manuais nos steps para dar uma ideia do potencial da ferramenta em termos de inserção de informações não sistematizadas. Vamos gerar a massa de dados do ano de 2003 a 2005 com o próprio PDI para esta dimensão.

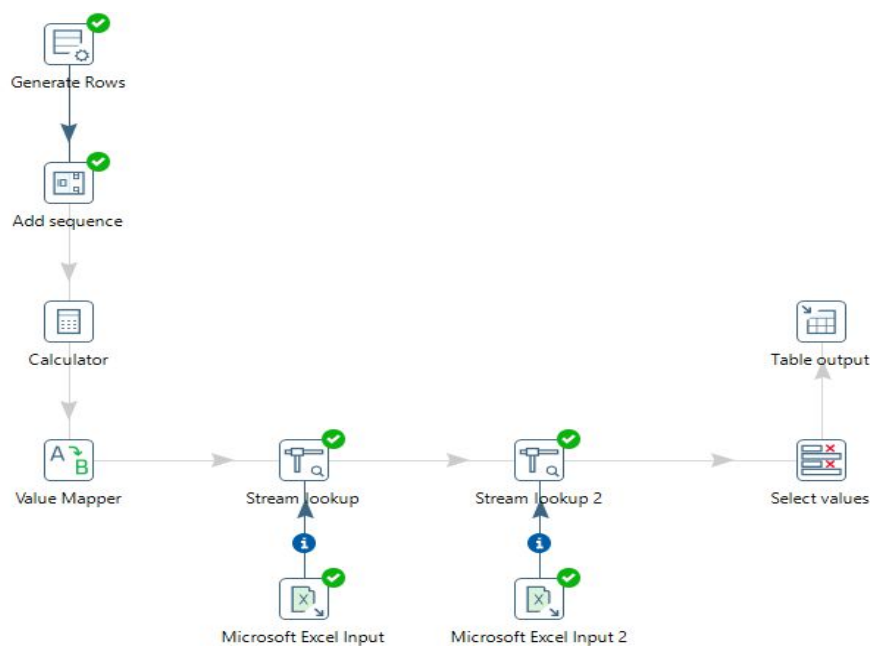


Figura 13-Transformação- Carregamento da dimensão Cliente no Staging Area

O fluxo se inicializa com o step Generated rows que tem como parâmetros a data de início e o limite de dias a gerar igual a 1090. Como resultado tem-se a coluna início com valores 20030106, que servirá de base para calcular as datas de acordo com o calendário. Na sequência temos o Add sequence, um contador, que gerará a coluna dia, disponibilizando assim possibilidades de consulta com o grão do dia. Estas últimas duas colunas são necessárias para a gerar as datas efetivas no step Calculator. Por ex. 20030106(início) + 26(dias) = 01/02/2003, respeitando assim o calendário. Para gerar os campos seguintes como ano, mes, diaano, diames, diasemana e semanaano, todos fazem referência ao campo data. Para gerar o trimestre vamos precisar do step Value mapper que mapeia o campo mês gerando o os trimestres. Caso seja necessário obter informações sobre valores de vendas mensal, a aplicação não disponibiliza esta informação em tipo texto, somente em números, então optou-se por trazer estas informações através de ficheiros de apoio em xls. Um com os dias da semana (DiaNr. DiaDesc, DiaDescAbrev) e outro com os meses (MesNr. MesDesc, MesDescAbrev). Consequentemente foram utilizados dois Stream lookups que busca os valores do dia da semana e os valores do dia do mês em String. Alguns atributos foram renomeados através do step Select value. Por fim Table output que faz o processo de carregamento de dados na dimensão tempo(tbl_dmns_tmp).

motor-inc-star.tbl_dmns_tmp: 27.351 registros totais (aproximadamente), limitado em 1.000

TempoId	DataTempo	AnoTempo	MesTempo	DiaAnoTempo	DiaMesTempo	DiaSemanaTempo	SemanaAnoTempo	DiaSemanaDescTempo	DiaSemanaDescAbrvTempo	MesDescTempo	MesDescAbrvTempo
1	2003-01-06 00:00:00	2.003	1	6	6	2	2	(NULL)	(NULL)	(NULL)	(NULL)
2	2003-01-07 00:00:00	2.003	1	7	7	3	2	(NULL)	(NULL)	(NULL)	(NULL)
3	2003-01-08 00:00:00	2.003	1	8	8	4	2	(NULL)	(NULL)	(NULL)	(NULL)
4	2003-01-09 00:00:00	2.003	1	9	9	5	2	(NULL)	(NULL)	(NULL)	(NULL)
5	2003-01-10 00:00:00	2.003	1	10	10	6	2	(NULL)	(NULL)	(NULL)	(NULL)
6	2003-01-11 00:00:00	2.003	1	11	11	7	2	(NULL)	(NULL)	(NULL)	(NULL)
7	2003-01-12 00:00:00	2.003	1	12	12	1	3	(NULL)	(NULL)	(NULL)	(NULL)
8	2003-01-13 00:00:00	2.003	1	13	13	2	3	(NULL)	(NULL)	(NULL)	(NULL)
9	2003-01-14 00:00:00	2.003	1	14	14	3	3	(NULL)	(NULL)	(NULL)	(NULL)
10	2003-01-15 00:00:00	2.003	1	15	15	4	3	(NULL)	(NULL)	(NULL)	(NULL)
11	2003-01-16 00:00:00	2.003	1	16	16	5	3	(NULL)	(NULL)	(NULL)	(NULL)
12	2003-01-17 00:00:00	2.003	1	17	17	6	3	(NULL)	(NULL)	(NULL)	(NULL)
13	2003-01-18 00:00:00	2.003	1	18	18	7	3	(NULL)	(NULL)	(NULL)	(NULL)
14	2003-01-19 00:00:00	2.003	1	19	19	1	4	(NULL)	(NULL)	(NULL)	(NULL)
15	2003-01-20 00:00:00	2.003	1	20	20	2	4	(NULL)	(NULL)	(NULL)	(NULL)
16	2003-01-21 00:00:00	2.003	1	21	21	3	4	(NULL)	(NULL)	(NULL)	(NULL)
17	2003-01-22 00:00:00	2.003	1	22	22	4	4	(NULL)	(NULL)	(NULL)	(NULL)
18	2003-01-23 00:00:00	2.003	1	23	23	5	4	(NULL)	(NULL)	(NULL)	(NULL)
19	2003-01-24 00:00:00	2.003	1	24	24	6	4	(NULL)	(NULL)	(NULL)	(NULL)
20	2003-01-25 00:00:00	2.003	1	25	25	7	4	(NULL)	(NULL)	(NULL)	(NULL)
21	2003-01-26 00:00:00	2.003	1	26	26	1	5	(NULL)	(NULL)	(NULL)	(NULL)
22	2003-01-27 00:00:00	2.003	1	27	27	2	5	(NULL)	(NULL)	(NULL)	(NULL)
23	2003-01-28 00:00:00	2.003	1	28	28	3	5	(NULL)	(NULL)	(NULL)	(NULL)
24	2003-01-29 00:00:00	2.003	1	29	29	4	5	(NULL)	(NULL)	(NULL)	(NULL)
25	2003-01-30 00:00:00	2.003	1	30	30	5	5	(NULL)	(NULL)	(NULL)	(NULL)
26	2003-01-31 00:00:00	2.003	1	31	31	6	5	(NULL)	(NULL)	(NULL)	(NULL)
27	2003-02-01 00:00:00	2.003	2	32	1	7	5	(NULL)	(NULL)	(NULL)	(NULL)

Figura 14-Tabela tempo na BD motor-in-star

6.7.1.5 Processo de carregamento da Dimensão Vendedor

A empresa Motor-inc possui seus colaboradores em seus cargos, como VP de Vendas, VP de Marketing, Gerente de Vendas e Representante de Vendas. Estas informações estão guardadas na coluna Cargo da tabela de origem empregados que se autorelacionam através da coluna SeReportaPara. E que se traduz na hierarquia da seguinte forma: O VP de Marketing, Jefferson e o VP de Vendas, Maria, possuem na coluna SeReportaPara o código do empregado Presidente, o seu superior direto, que é Diana. Por vezes a empresa precisa de levantar os dados de quanto cada vendedor vendeu, o quanto cada gerente vendeu, ou seja, a somatização de vendas de todos seus vendedores e a maneira que os dados estão guardados, não é possível fazer este levantamento. A complementação da ideia também é ir buscar através do território esta informação pois cada lugar que o vendedor esta tem uma regra de participação dos lucros, então se um vendedor é transferido de uma região para outra esta informação deve ser guardada para que possa aplicar as regras de participação dos lucros de acordo com o localização , para isso é necessário aplicar o Slow Changing Dimension do tipo 2. E é o que vamos detalhar nesta transformação.

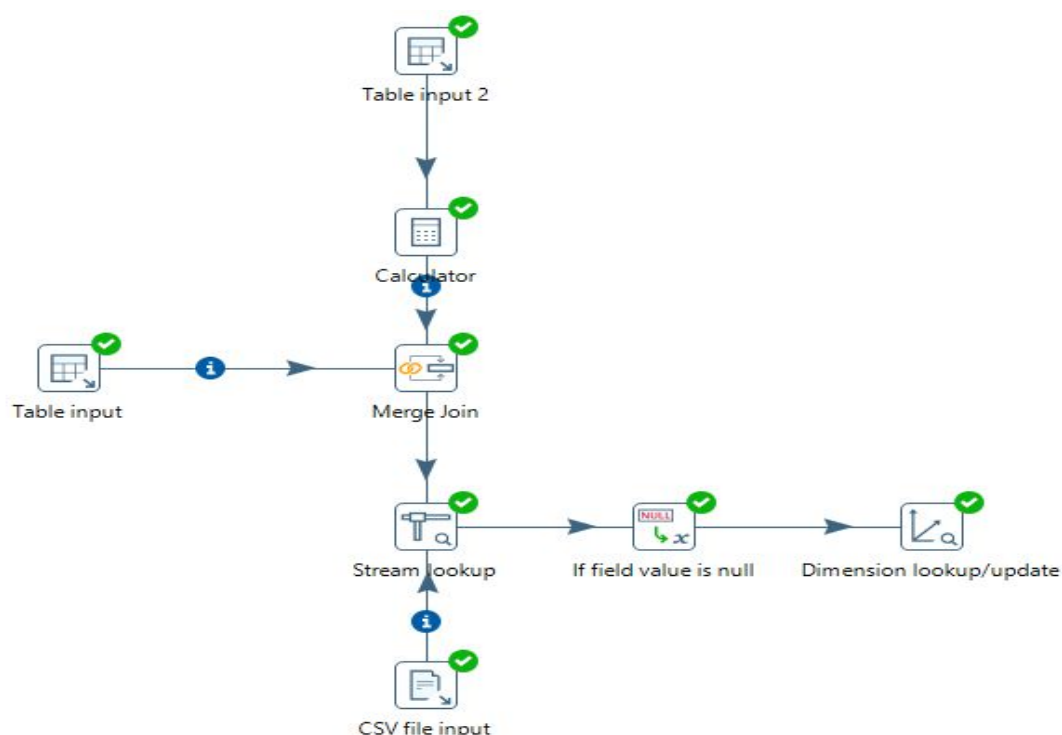


Figura 15-Transformação- Carregamento da dimensão Vendedor no Staging Area

O fluxo inicializa-se com um Table input que traz uma Query na desnormalização dos três níveis de hierarquia cargo (Presidente, VP Marketing e Gerente de Vendas). Esta desnormalização traz os nomes dos vendedores relacionado a cada um de seu gerente. O Table input 2 trás uma Query com o resultado de todos os cargos, menos o de Presidente. Os nomes e sobrenomes são concatenados no step Calculator, seguidamente temos o step Merge join (LEFT OUTER), que traz todos cargos mesmos os que não possuem cargo superior relacionado na hierarquia, como o VP de Marketing que não possui gerente. Compara-se o valor do NumeroEmpregado1 da tabela 1 ao valor da coluna seReportaPara do Calculator. Todos os números dos vendedores que forem iguais aos códigos dos Gerentes, é porque é vendedor daquele Gerente. O Merge join precisa de receber os dados já ordenados, para que seja feita as comparações entre as chaves passadas por parâmetros, que já foram definidos nas queries das Tables input. O Stream lookup vai consumir os dados complementares como cidade, estado, país e território de cada vendedor que dá uma ideia de Staging também, porém uma Staging em ficheiro. Como nesta tabela existem valores nulos e queremos substituir estes valores por uma informação mais relevante utilizamos o If valeu is Null, inserindo para todos valores 'NULL' o valor em String 'SEM INFORMAÇÃO'. Como o VP de Marketing que não possuem gerentes.

Query Table input 1.

```
SELECT
    A.NumeroEmpregado AS NumeroEmpregado3
    , CONCAT (CONCAT (A.PrimeiroNome, ' '), A.UltimoNome) AS NomeVendedor3
    , B.NumeroEmpregado AS NumeroEmpregado2
    , CONCAT (CONCAT (B.PrimeiroNome, ' '), B.UltimoNome) AS NomeVendedor2
    , C.NumeroEmpregado AS NumeroEmpregado1
    , CONCAT (CONCAT (C.PrimeiroNome, ' '), C.UltimoNome) AS NomeVendedor1
FROM empregados A
INNER JOIN empregados B ON A.NumeroEmpregado = B.SeReportaPara
LEFT OUTER JOIN empregados C ON B.NumeroEmpregado = C.SeReportaPara
WHERE A.SeReportaPara IS NULL

order by 5
```

Query Table input 2

```
SELECT
    NumeroEmpregado
    , UltimoNome
    , PrimeiroNome
    , Ramal
    , CorreioEletronico
    , CodigoEscritorio
    , SeReportaPara
    , Cargo
FROM empregados
WHERE SeReportaPara is not null

ORDER BY SeReportaPara
```

Toda tabela para que seja aplicado o conceito de SCD tipo 2 deve possuir uma estrutura de colunas de controle como data de início, data fim e data da inserção do registro, pois são nestas colunas que são guardados os dados de histórico. Este procedimento foi satisfeito na transformação com o step Dimension Lookup/Update, fazendo a comparação do NumeroDoEmpregado da tabela vendedor e NumeroDoEmpregado que vem do fluxo (tabela empregado). Neste step temos que definir a chave natural da tabela empregado (BD motor-inc-oltp), a data início da vigência, a data final e do campo Versão para gravar a versão do registro. Este step é utilizado aqui no contexto Update e faz duas verificações: a primeira é após a comparação entre as chaves NumeroDoEmpregado caso não exista, ele insere, caso exista ele faz uma segunda verificação. Verifica-se o valor do campo território se são os mesmos, caso não houve mudança de valor nesta coluna, então não se aplica nada, caso houve mudança de valor então faz Inserção (Insert) dos novos valores do registro (IdVendedor, paísVendedor, territórioVendedor, data_inicial, data_final). Assim é aplicado o Slow Changing Dimension que guarda histórico da informação. Para o restante dos campos faz um Update dos valores, caso precise, perdendo assim os valores antigos, pois o SCD não é necessário ser aplicado em todos os campos, mas sim nas informações impactantes no negócio, como o território.

motor-inc-star.tbl_dmns_vnddr: 52 registros totais (aproximadamente)										Próximo	Mostrar todos	Ordem	Colunas (16/16)	Filtro
Vend...	Numero...	Nome/Vendedor3	Numero...	Nome/Vendedor2	Numero...	Nome/Vendedor1	Numero...	Nome/Vendedor	Cidade/Vendedor	Estado/Vendedor	Pais/Vendedor	Territorio...	versao	data_inicial
(NULL)	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)	1	(NULL)
1	1.002	Diana Murphy	1.076	Jefferson Firrell	-1	SEM INFORMA...	-1	SEM INFORMA...	SEM INFORMA...	SEM INFORMA...	SEM INFORMA...	SEM INF...	1	1900-01-01 00:00:00
2	1.002	Diana Murphy	1.056	Maria Patterson	1.088	William Patterson	1.611	André Fixter	SIDNEI	SEM INFORMA...	AUSTRALIA	APAC	1	1900-01-01 00:00:00
3	1.002	Diana Murphy	1.056	Maria Patterson	1.088	William Patterson	1.612	Pedro Marsh	SIDNEI	SEM INFORMA...	AUSTRALIA	APAC	1	1900-01-01 00:00:00
4	1.002	Diana Murphy	1.056	Maria Patterson	1.088	William Patterson	1.619	Tomas King	SIDNEI	SEM INFORMA...	AUSTRALIA	APAC	1	1900-01-01 00:00:00
5	1.002	Diana Murphy	1.056	Maria Patterson	1.088	William Patterson	1.621	Mami Nishi	TOQUIO	CHITYODA-KU	JAPÃO	JAPÃO	1	1900-01-01 00:00:00
6	1.002	Diana Murphy	1.056	Maria Patterson	1.088	William Patterson	1.625	Yoshimi Kato	TOQUIO	CHITYODA-KU	JAPÃO	JAPÃO	1	1900-01-01 00:00:00
7	1.002	Diana Murphy	1.056	Maria Patterson	1.102	Geraldo Bondur	1.504	Bruno Jones	LONDRES	SEM INFORMA...	REINO UNIDO	EMEA	1	1900-01-01 00:00:00
8	1.002	Diana Murphy	1.056	Maria Patterson	1.102	Geraldo Bondur	1.501	Larry Bott	LONDRES	SEM INFORMA...	REINO UNIDO	EMEA	1	1900-01-01 00:00:00
9	1.002	Diana Murphy	1.056	Maria Patterson	1.102	Geraldo Bondur	1.401	Pamela Castillo	PARIS	SEM INFORMA...	FRANÇA	EMEA	1	1900-01-01 00:00:00
10	1.002	Diana Murphy	1.056	Maria Patterson	1.102	Geraldo Bondur	1.370	Geraldo Herman...	PARIS	SEM INFORMA...	FRANÇA	EMEA	1	1900-01-01 00:00:00
11	1.002	Diana Murphy	1.056	Maria Patterson	1.102	Geraldo Bondur	1.337	Loui Bondur	PARIS	SEM INFORMA...	FRANÇA	EMEA	1	1900-01-01 00:00:00
12	1.002	Diana Murphy	1.056	Maria Patterson	1.102	Geraldo Bondur	1.702	Marcos Gerard	PARIS	SEM INFORMA...	FRANÇA	EMEA	1	1900-01-01 00:00:00
13	1.002	Diana Murphy	1.056	Maria Patterson	1.143	Antonio Bow	1.323	Jorge Vanauf	SÃO FRANCISCO	CA	ESTADOS UNID...	NA	1	1900-01-01 00:00:00
14	1.002	Diana Murphy	1.056	Maria Patterson	1.143	Antonio Bow	1.286	Foon Yue Tseng	BOSTON	MA	ESTADOS UNID...	NA	1	1900-01-01 00:00:00
15	1.002	Diana Murphy	1.056	Maria Patterson	1.143	Antonio Bow	1.216	Estevan Patterson	BOSTON	MA	ESTADOS UNID...	NA	1	1900-01-01 00:00:00
16	1.002	Diana Murphy	1.056	Maria Patterson	1.143	Antonio Bow	1.188	Julia Firrell	BOSTON	MA	ESTADOS UNID...	NA	1	1900-01-01 00:00:00
17	1.002	Diana Murphy	1.056	Maria Patterson	1.143	Antonio Bow	1.166	Lella Thompson	SÃO FRANCISCO	CA	ESTADOS UNID...	NA	1	1900-01-01 00:00:00
18	1.002	Diana Murphy	1.056	Maria Patterson	1.143	Antonio Bow	1.165	Lella Jennings	SÃO FRANCISCO	CA	ESTADOS UNID...	NA	1	1900-01-01 00:00:00
19	1.002	Diana Murphy	1.056	Maria Patterson	1.143	Antonio Bow								

Figura 16-Tabela vendedor na BD motor-in-star.

6.7.1.6 Processo do Carregamento da Tabela de Facto

O objetivo desta transformação é apresentar o processo de carga da tabela fato Vendas. Nesta transformação, além das medidas que devem ser apresentadas (Quantidade e Valor Total), vamos trazer também as chaves primárias, para o processo do surrogate Key na tabela de facto. A carga de dados é sempre inserção e há um tratamento de erro de violação de constraint, caso ocorra, o registo será expurgado para um arquivo texto. Neste momento iremos também gerar uma informação nova, um procedimento de enriquecimento de dados.

A empresa Motor-inc não possui informações de vendas de seus colaboradores por região, pois as vendas estão todas guardadas na coluna. Caso um vendedor mude de um escritório para outro, as informações sobre suas vendas eram perdidas. Até se aplicar o SCD tipo2 na Tabela vendedor, e agora vamos trazer estas informações aqui.

Para que possamos obter informação de onde foram feitos as Vendas, vamos trabalhar a busca das datas das vendas na tabela cliente para saber qual intervalo de vendas estão associados aos respectivos escritórios.

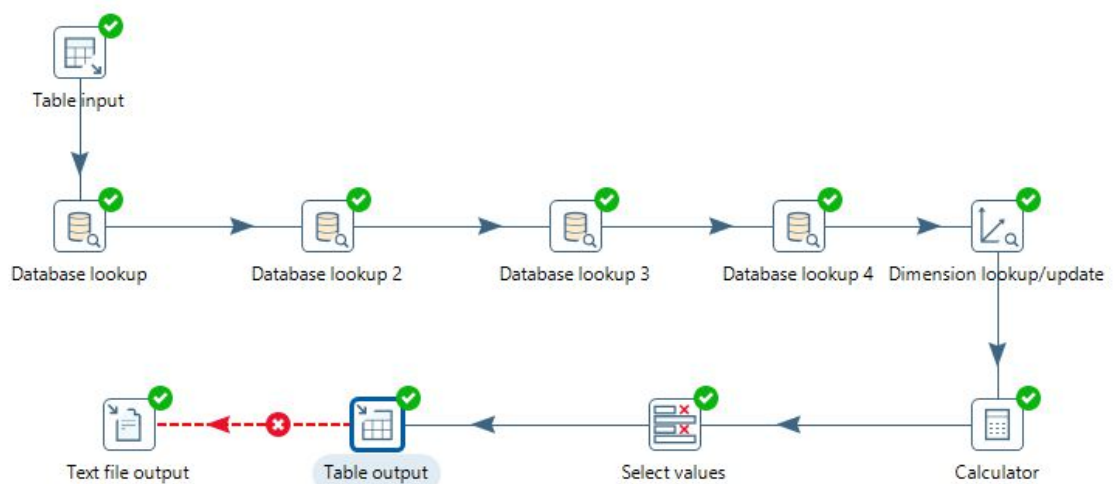


Figura 17Transformação- Carregamento da tabela de facto.

No Processo de extração das Surrogates Keys para cada dimensão foram utilizados lookups para fazer a comparação entre as chaves naturais vindas das tabelas de origem, com as chaves conhecidas como DD (Dimensão Degenerada) inseridas nas dimensões. Como resultado a função retorna a chave técnica (chave primaria) da dimensão em questão. Este procedimento se aplicará para todos os lookUps do fluxo para que a tabela de facto possa trabalhar com estas chaves que representam as surrogates Keys.

Para a dimensão tempo foram utilizados dois Database lookup, pois a data do pedido não é a mesma que a data de entrega, e assim aplica-se o conceito do Role Playing Dimension, no caso o tempo pedido e o tempo entrega. No Dimension LookUp Update, quando a facto é gerada, o fluxo permite pegar a data do pedido através do NumeroEmpregadoVendedor da tabela cliente da origem com o NumeroEmpregado da dimensão vendedor, assim a função trás as informações de vendas respetivas aos seus escritórios. O step Calculator vai trazer a medida do valor total de vendas e no Select values os campos que serão mantidos no fluxo acrescido do atributo Situação que é esta sendo representado no conceito de uma dimensão degenerada para o tratamento de análise Olap.

Para finalizar temos o Table output que carrega a tabela de facto e a Text File Output que configura o error handling com uma funcionalidade de gestão de erros que pode identificar qual o número do erro gerado, em qual coluna, etc. Para qualquer tipo de violação de constraint, estes erros serão expurgados para este ficheiro.

Query da Table input

```
SELECT
    detalhespedidos.NumeroPedido
,   detalhespedidos.CodigoProduto
,   detalhespedidos.QuantidadePedida
,   detalhespedidos.PrecoUnitario
,   detalhespedidos.NumeroLinhaPedido
,   pedidos.DataPedido
,   pedidos.DataEntrega
,   pedidos.Situacao
,   pedidos.NumeroCliente
,   clientes.NumeroEmpregadoVendedor
FROM detalhespedidos
INNER JOIN pedidos
ON detalhespedidos.NumeroPedido = pedidos.NumeroPedido
INNER JOIN clientes
ON pedidos.NumeroCliente = clientes.NumeroCliente
```

motor-inc-olap-front1000.tbl_ft_vnds: 2.854 registros totais (aproximadamente), limitado em 1.000

Próximo Mostrar todos Ordem

id	NumeroPedido	NumeroLinhaPedido	Situacao	TempoIdPedido	TempoIdEntrega	ProdutoId	ClienteId	VendedorId	QuantidadePedida	PrecoTotal
1	10.100	1	E	1	5	90	36	15	49	1.666
2	10.100	2	E	1	5	92	36	15	50	3.400
3	10.100	3	E	1	5	97	36	15	30	5.160
4	10.100	4	E	1	5	104	36	15	22	1.914
5	10.101	1	E	4	6	106	54	7	26	3.770
6	10.101	2	E	4	6	87	54	7	46	2.484
7	10.101	3	E	4	6	93	54	7	45	1.395
8	10.101	4	E	4	6	95	54	7	25	3.775
9	10.102	1	E	5	9	99	66	14	41	2.050
10	10.102	2	E	5	9	91	66	14	39	4.797
11	10.103	1	E	24	28	79	52	7	36	3.672
12	10.103	2	E	24	28	84	52	7	22	1.188
13	10.103	3	E	24	28	78	52	7	31	3.224
14	10.103	4	E	24	28	50	52	7	42	5.418
15	10.103	5	E	24	28	77	52	7	36	4.212
16	10.103	6	E	24	28	73	52	7	42	4.452
17	10.103	7	E	24	28	76	52	7	45	3.420
18	10.103	8	E	24	28	80	52	7	27	3.402
19	10.103	9	E	24	28	100	52	7	41	1.927
20	10.103	10	E	24	28	82	52	7	35	3.920
21	10.103	11	E	24	28	39	52	7	26	5.408
22	10.103	12	E	24	28	107	52	7	27	2.241
23	10.103	13	E	24	28	89	52	7	25	2.550
24	10.103	14	E	24	28	85	52	7	35	1.995
25	10.103	15	E	24	28	94	52	7	25	2.875
26	10.103	16	E	24	28	102	52	7	46	4.784
27	10.104	1	E	26	27	55	55	10	34	5.950
28	10.104	2	E	26	27	109	55	10	32	1.696

Figura 18-Carregamento da tabela de facto Vendas na BD motor-in-star

P.S. Neste Trabalho não será representada a transformação que faz o carregamento da dimensão Produto uma vez que estes dados já estão disponíveis na Base de dados.

6.7.1.7 Jobs- Orquestração da execução dos processos

As Transformações servem para a movimentação dos dados, e os jobs servem para a orquestração da execução dos processos. O procedimento aqui vai utilizar as dimensões que já foram carregadas, por isso não serão detalhados os passos dos jobs, uma vez que já estão todos definidos no capítulo anterior.

Processo de execução de um job ao carregar as tabelas território e cliente_territorio para a área de estagiamento.

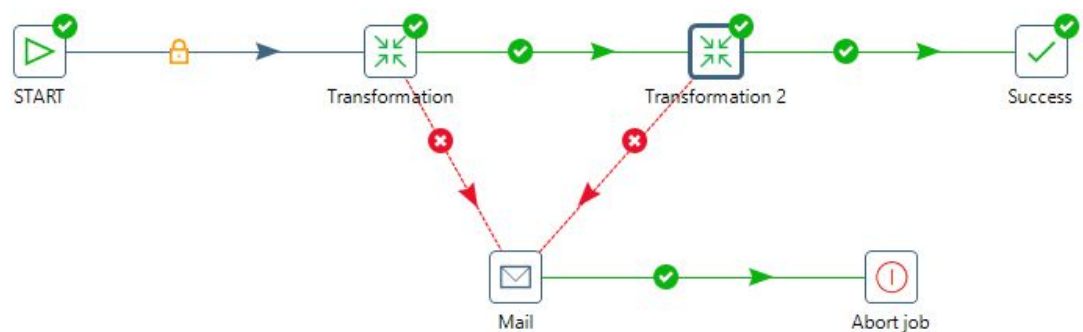


Figura 19-Execução do job no processo de transformações de território e cliente_territorio.

Neste Job, apresentaremos como encadear as chamadas às transformações que carrega as tabelas no Staging Area. O processo pode resultar em sucesso ou em falha que ocasiona em uma ação de envio de e-mail. A execução das transformações devem ser uma seguida da outra. O step Start inicia a execução do job, seguidamente procede-se uma transformação que é executada (tabela território), caso a execução do processo seja positiva, então segue para a próxima transformação (carregamento tabela carga_stage_cliente_territorio) e segue a sequência positiva executando a mensagem de sucesso. Caso o resultado seja falso, segue a hope com handling error já configurada para o próximo step que despoleta o envio de um e-mail. Mas o processo pode ter algum problema ao enviar o e-mail, por ex. ele não acha o servidor de e-mail, isso acarreta uma falha, então será configurado como incondicional na hope entre e-mail e o step Abort job. Para que o job seja abortado, independente dele ser enviado com sucesso ou não.

Processo de execução de um job ao carregar as dimensões Vendedor e Cliente para a área de estagiamento.

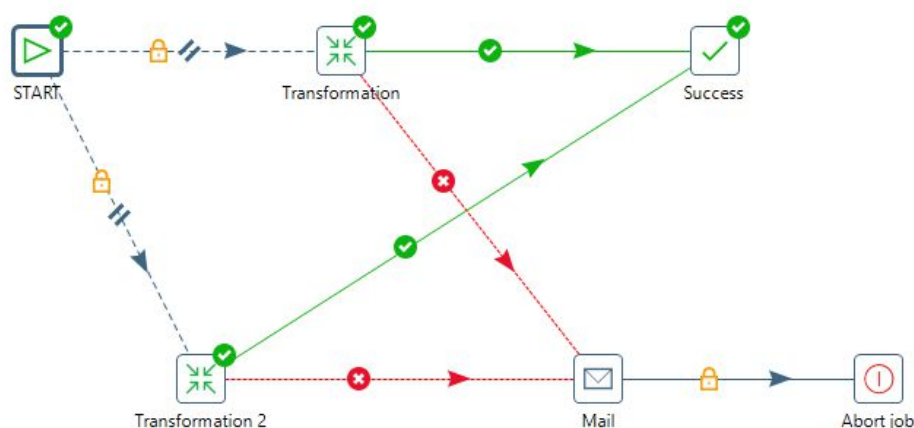


Figura 20-Execução do job no processo de transformações de Vendedor e Cliente.

Este job tem a finalidade de encadear as chamadas pelos processos que carregam as dimensões de Cliente e Vendedor, resultando em sucesso ou em falha que aciona o envio de um e-mail de erro. Neste contexto as duas transformações são executadas em paralelo, pois nenhuma depende da outra. As tabelas são cliente e vendedor.

Para que as hopes(ligações) sejam executadas ao mesmo tempo no outbound do step Start, é necessário configurar a opção Run Next Entries in Parallel. Assim o processo garante que as duas hopes são executadas ao mesmo tempo. Em seguida o procedimento é o mesmo que foi executado no job anterior.

Uma outra observação neste processo é que a tabela de facto só acionará a ação de execução quando as duas dimensões forem carregadas com sucesso. Um dos motivos que a tabela de facto não esta neste job.

Construção do job na execução da tabela Facto

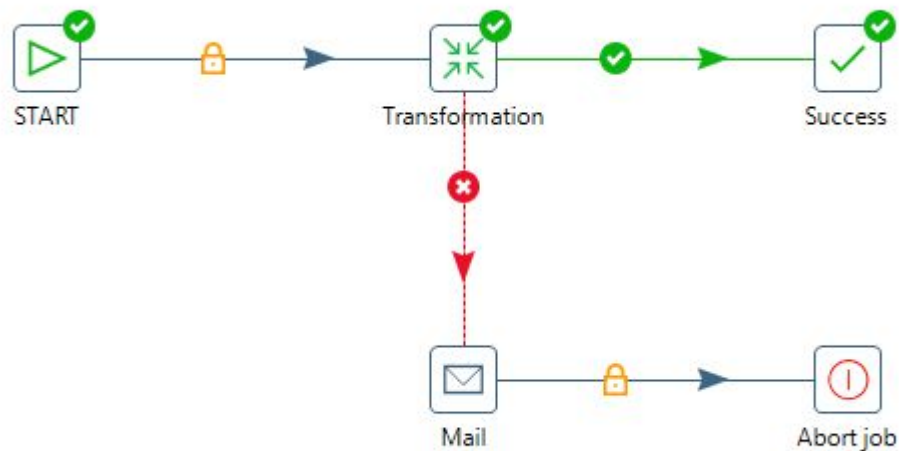


Figura 21-Execução do job no processo da transformação da tabela de facto.

Neste Job, apresentaremos como encadear a chamada ao processo da tabela fato vendas, resultando em sucesso ou em falha e neste caso, enviando e-mail de erro.

Mesmo esquema dos outros dois acima, se executou com sucesso, gera uma mensagem de sucesso, se falhar envia o email, e independente se conseguiu enviar o email ou não, aborta todo o processo no step Abort job.

Orquestrando o processo do Data wareHouse

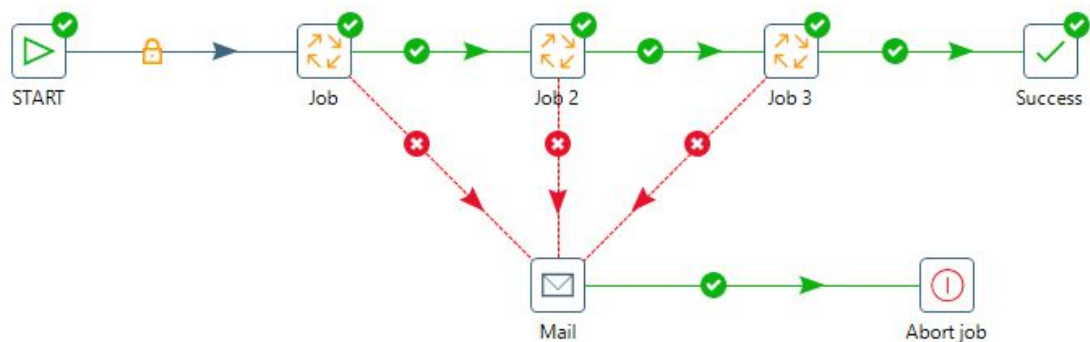


Figura 22-Execução do processo vários jobs para o carregamento no DW.

Este Job demonstra o encadeamento das chamadas na execução do job 1 , que representa o carregamento das área de estagiamento, na sequência a execução do job 2, configurando o carregamento das dimensões e por fim o processo do job 3, que representa o carregamento da

tabela de fato vendas, se o resultado for de sucesso, uma mensagem é enviada, caso o resultado seja de falha, um e-mail será enviando.

6.8 Análises dos Dados em Cubos Olap

Análises dos dados é uma característica essencial entre as soluções de BI, tende a ser mais dinâmica que os relatórios e é usada pelos gestores para a tomada de decisões ao nível estratégico. Através da análise ad-hoc, o utilizador pode explorar os dados dinamicamente, normalmente apresentados de em um nível mais agregado, por exemplo, as vendas totais de produtos por ano e ir navegando por um nível mais detalhado, como as vendas por mês e região. O modelo dimensional que está intimamente relacionado às análises dos dados é que permite a visualização e aplicação destes métodos de agregação e de navegação Drill down e Roll Up utilizando ferramentas adequadas para tal. A ferramenta Mondrian é utilizada para a exploração dos dados através de ficheiros XML nas definições das dimensões, factos e suas hierarquias na condição da criação dos cubos OLAP no Schema Workbench e de sua montagem e visualização no Saiku (Lima, 2012).

A aplicação Schema Workbench é uma interface gráfica que permite criar e testar cubos Olap no Mondrian. Utilizada neste projeto para demonstrar a criação dos cubos que atendem às navegações por medição da quantidade de produtos vendidos, valor total de vendas e seu percentual. A definição de um cubo é constituída por uma tabela de facto, por suas dimensões, os níveis de hierarquia e medidas (Lima, 2012). Cada cubo deve ser baseado somente numa tabela de facto. As medidas representam as colunas provenientes da tabela de factos que contém os dados numéricos que se pretendem analisar. (Dongen, 2009).

6.8.1 Estrutura dos Cubos

- **Schema: Motor Inc.**
 - **Cubos:** Motor inc. Vendas; VendasOrca, Motor-Inc.-Metas
 - **Medidas:** Quantidade Pedida, Preço Total, percentual do Produto
 - **Dimensões:** Cliente, Produto, Tempo, Vendedor
 - **Hierarquias**
 - Tempo:
 - Calendário: Ano, Trimestre; Mês
 - Semana: Ano Semana
 -
 - Cliente: Território, País, Estado, Cliente
 - Produto: Linha, Fornecedor, Produto
 - Vendedor: Presidente, Gerente, VP, Vendedor.
 - Níveis - Atributos

6.8.2 Tabelas agregadas

Alguns fatores podem interferir negativamente no desempenho das consultas às Bases de dados como a quantidade de volume dos dados, a forma em que o Modelo dimensional foi implementado e a própria definição do grão. Porém o Pentaho oferece como solução as tabelas agregadas para o melhor desempenho das análises em cubos Olap.

Como já foi identificado anteriormente, existe uma hierarquia com um número de níveis que deve ser seguido para a montagem do cubo. A tabela `a_t_motor_inc` da BD `motor-inc-olap` é uma tabela sumarizada, e que contém uma coluna denominada `factcount` com valores 1 que faz a identificação das tabelas agregadas na BD, esta coluna é essencial para identificação da tabela agregada alvo. O Mondrian que é o motor Olap no BI Server, utiliza consultas MDX na BD para verificar se existem tabelas agregadas. Vamos demonstrar em um exemplo fictício da seguinte forma: Caso existam 10 tabelas agregadas, mas só 3 atende a requisição, o Mondrian faz um count na coluna `factcount` relacionadas a estas três tabelas e identifica qual delas retornar o menor número de registros, uma vez identificada a tabela alvo, é feito o parse do mdx para o sql. Se a tabela de facto não possui tabelas agregadas, ou se alguma das tabelas agregadas existentes não atende a nenhuma das requisições feitas, então o Mondrian faz o parse do MDX para o sql direcionado na tabela de facto. Logo abaixo na Figura 23 que mostra a adição da tabela agregada `a_t_motor_inc` no Schema Workbench, que demonstra os principais passos para sua configuração.

A primeira coisa a se fazer antes de criar uma estrutura das tabelas agregadas no cubo é configurar dois ficheiros para que as tabelas possam ser executadas. A primeira configuração que deve ser feita está relacionada ao Mondrian, pois ele não reconhece por padrão se tem tabelas agregadas. Esta configuração pode-se ver no Apendice 1 do item 2, Configuração do ficheiro `properties` no Mondrian. A segunda configuração a fazer permite o utilizador acompanhar os logs do Mondrian na BD, e refere-se ao item 3 do Apendice 1, denominado Configuração de log no Mondrian.

Nós sabemos que podemos ter dois tipos de tabelas agregadas. O conceito de tabelão, onde são agregados todos os dados, onde a query é feita em 100% dos dados que estão nesta tabela, e caso a pesquisa não é atendida, o Mondrian volta para a tabela de Facto. Este caso é considerado do primeiro meio e é mais indicado para fazer consultas para Dashboards, pois as consultas já estão pre-definidas e são mais rápidas para este output. Temos também as tabelas agregadas do segundo meio, que estão preparadas para receber consultas mais dinâmicas, consultas Ad-hoc. Essas tabelas agregadas possuem relações com as dimensões que guardam seus IDs, e caso a pesquisa não for atendida 100%, pode-se fazer a análise através destes joins com as dimensões, e mesmo assim as consultas ainda ficam mais rápidas do que retornar a tabela de Facto e refazer as consultas. O exemplo a seguir será demonstrado no conceito tipo 2 de agregação.

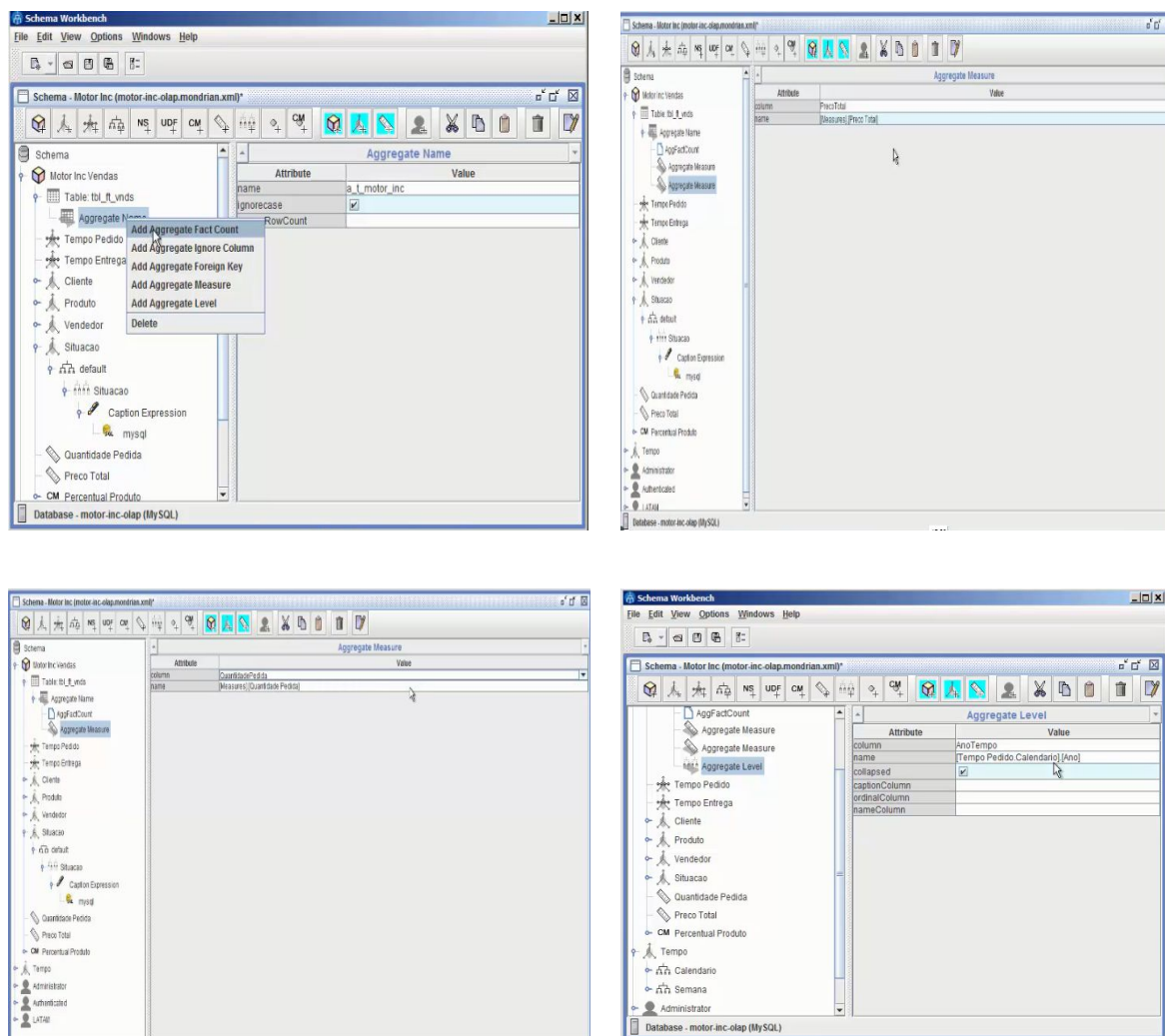


Figura 23- Definição da tabela Agregada a_t_motor_inc 1ª parte.

No primeiro quadrante temos o Mapeamento do campo Factcount da tabela agregada at_motor_inc. Este campo faz a identificação das tabelas agregadas na BD para a tabela de facto, como já foi mencionado acima e deve ser inserida logo abaixo na hierarquia da tabela de Facto tbl_ft_vnds. No objeto Aggregate Name, adicionamos o objeto Aggregate Fact Count e definimos a propriedade column como factcount.

O segundo e terceiro quadrante mostra a adição das medidas, o preço total e a quantidade pedida. Ainda no objeto Aggregate Name, adicionamos o objeto Aggregate Measure e definimos as propriedades column como PrecoTotal e QuantidadePedida, e para name como [Measures].[Preco Total] e [Measures].[Quantidade Pedida], respectivamente;

E para finalizar no 4º quadrante, a dimensão do Tempo pedido, uma vez que queremos obter as datas que foram feitas os pedidos das vendas.

adicionamos o objeto Aggregate Level e definir as propriedades column como AnoTempo e name como [Tempo Pedido.Calendario]. [Ano];

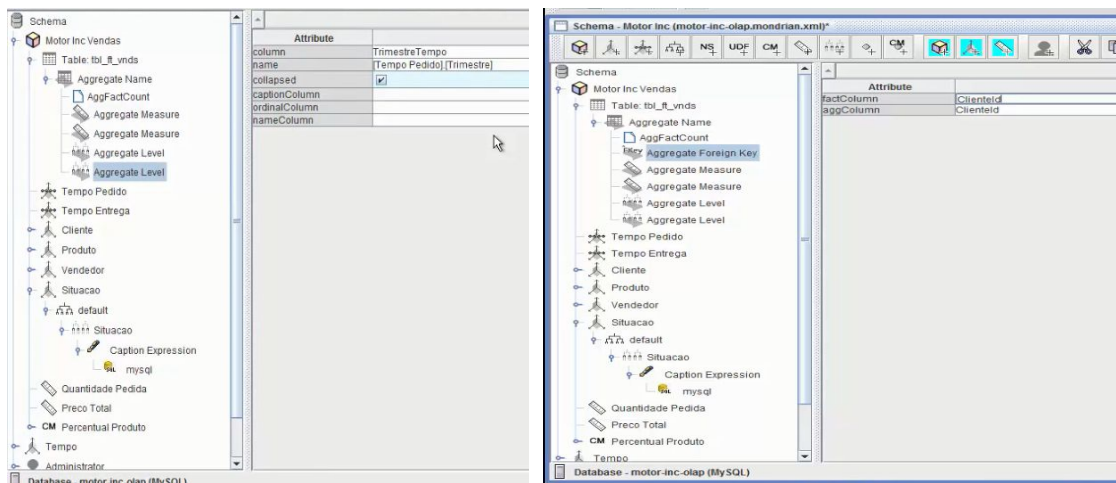


Figura 24- Definição da tabela Agregada a_t_motor_inc 2ª parte.

Na primeira tabela da figura 24 temos a configuração da adição do trimestre para que a consulta possa ir num nível mais baixo do grão que ficou definido no objeto Aggregate Level as propriedades column como TrimestreTempo e name como [Tempo Pedido.Calendario].[Trimestre]. E na sequência a adição da primeira chave estrangeira, a surrogate key Cliente da tabela agregada. Sempre no objeto Aggregate Name, adicionamos o objeto Aggregate Foreign Key e definimos as propriedades factColumn e aggColumn respectivamente como Clienteld. O mesmo procedimento foi feito para as chaves estrangeiras da dimensão produto e vendedor.

Um exemplo a ser aplicado na requisição à tabela agregada, seria uma consulta à quantidade total de produtos e preço de vendas por ano. Neste caso o Mondrian faria a consulta na tabela agregada relacionado ao ano, e faz um innerjoin na dimensão Produto através da Surrogatekey idProduto que se encontra na tabela agregada. Caso fosse uma requisição num nível mais granular com a quantidade total de produtos e preço de vendas por mês, o Mondrian faria a pesquisa na tabela de facto, pois não existe este atributo na tabela agregada e não existe também uma relação entre a tabela agregada com a dimensão Tempo para consultar o atributo mês. Neste caso o Mondrian manda a query para a tabela Facto.

A tabela agregada pode ser visualizada logo abaixo na Figura 25.

AnoTempo	TrimestreTempo	🌱 ProdutoId	🌱 ClienteId	🌱 VendedorId	QuantidadePedida	PrecoTotal	factcount
2.003	TRIM1	1	81	9	31	3.317	1
2.003	TRIM1	2	81	9	28	2.492	1
2.003	TRIM1	3	81	9	32	4.000	1
2.003	TRIM1	4	81	9	36	5.292	1
2.003	TRIM1	5	81	9	34	3.774	1
2.003	TRIM1	6	81	9	44	3.256	1
2.003	TRIM1	7	81	9	48	2.544	1
2.003	TRIM1	8	81	9	50	3.250	1
2.003	TRIM1	9	81	9	48	2.928	1
2.003	TRIM1	10	81	9	49	3.675	1
2.003	TRIM1	11	81	9	26	1.664	1
2.003	TRIM1	12	81	9	34	3.060	1
2.003	TRIM1	13	81	9	41	3.403	1
2.003	TRIM1	14	56	9	31	2.046	1
2.003	TRIM1	15	56	9	41	2.911	1
2.003	TRIM1	16	56	9	22	2.552	1
2.003	TRIM1	17	81	9	30	3.180	1
2.003	TRIM1	18	56	9	25	1.425	1
2.003	TRIM1	19	56	9	29	2.030	1
2.003	TRIM1	20	56	9	43	6.321	1
2.003	TRIM1	21	56	9	39	3.159	1
2.003	TRIM1	23	6	13	30	2.880	1
2.003	TRIM1	24	6	13	39	3.900	1
2.003	TRIM1	25	6	13	38	3.154	1
2.003	TRIM1	26	106	5	35	2.065	1
2.003	TRIM1	27	6	13	27	6.075	1
2.003	TRIM1	28	106	5	31	3.658	1
2.003	TRIM1	29	6	13	25	2.850	1

Figura 25- Tabela agregada a_t_motor_inc.

6.8.3 Role Playing Dimension e a Dimensão Tempo

A dimensão Tempo é considerada uma dimensão em conformidade pois pode ser utilizada em vários cubos, ou até mesmo fora dele. Vamos considerar para o primeiro caso a dimensão Tempo criada fora do cubo com as hierárquias Calendário e seus atributos Ano, Trimestre e Mês, e a hierarquia Semana com seus atributos Ano e Semana do Ano. Assim o modelo disponibiliza possibilidades mais dinâmicas de navegação, caso o utilizador precise de fazer consultas mais pontuais. Os níveis hierárquicos estabelecidos permitirão o utilizador a navegar através de recursos Drill-down e Roll-up.

Para o segundo caso, onde a dimensão Tempo foi criada dentro do cubo, aplicou-se o conceito de Role Palyng Dimension que faz referência para as outras duas dimensões Tempo do pedido e Tempo da entrega sem precisar de fazer o mapeamento 2 vezes. A Dimension Usage é o recurso criado para este procedimento como é mostrado na figura abaixo.

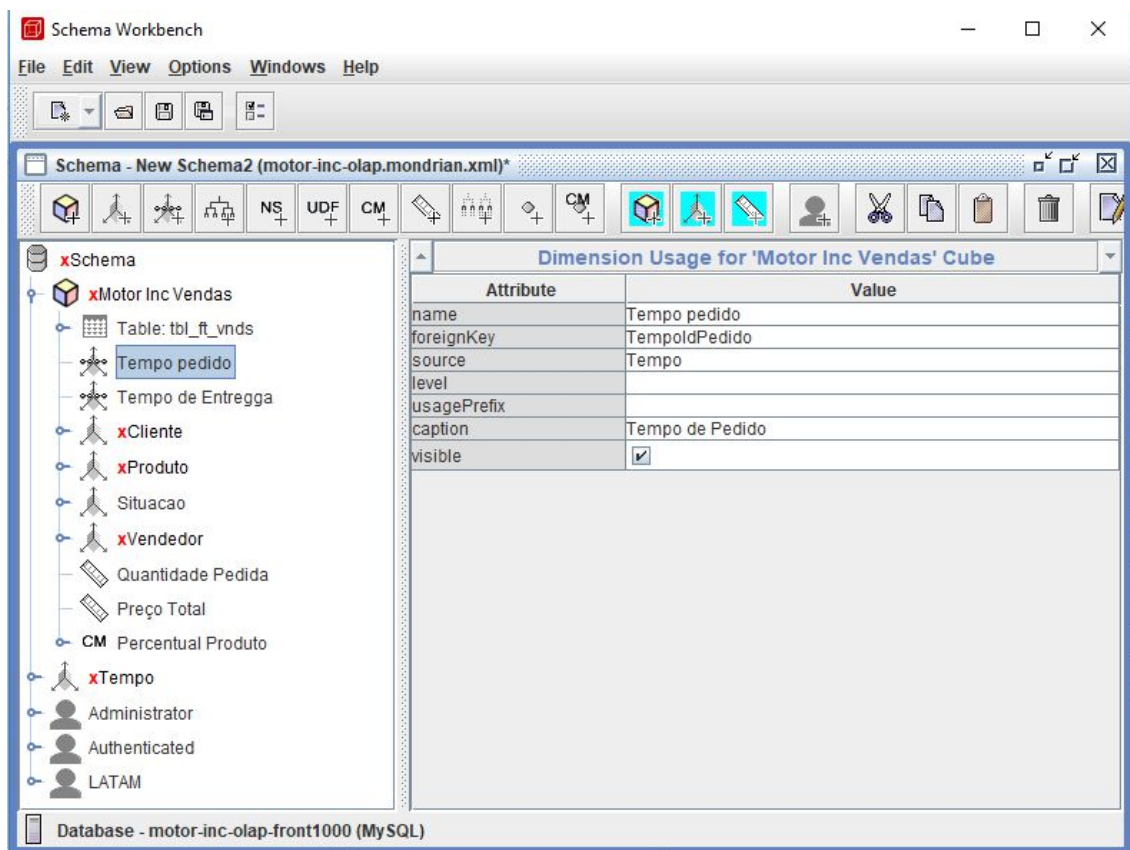


Figura 26-Configuração da dimensão tempo no conceito de Role Playng Dimension.

6.8.4 Dimensão Degenerada

Na tabela de Facto tb_ft_vnds existe um campo denominado Situação que reflete no status de triagem dos produtos vendidos. Como enviado, cancelado, Resolvido, etc. Este atributo é representado nesta tabela de facto no conceito Degeneration Dimension ou Dimensão Degenerada, que possui o conceito lógico de uma tabela apesar de não existir uma tabela física na DB. O mapeamento desta dimensão para o utilizador é de total transparência e seus dados podem ser acessados via PUC . A definição de cada status na tabela é representada por um caractere pois quanto maior a optimização dos dados em uma tabela de facto, melhor seu desempenho. Como pode ser visto na figura 27, cada letra é definida para ser mais representativa para o utilizador final e que ficou da seguinte forma:E- enviado, R- Resolvido, A- Aguardando e C- Cancelado.

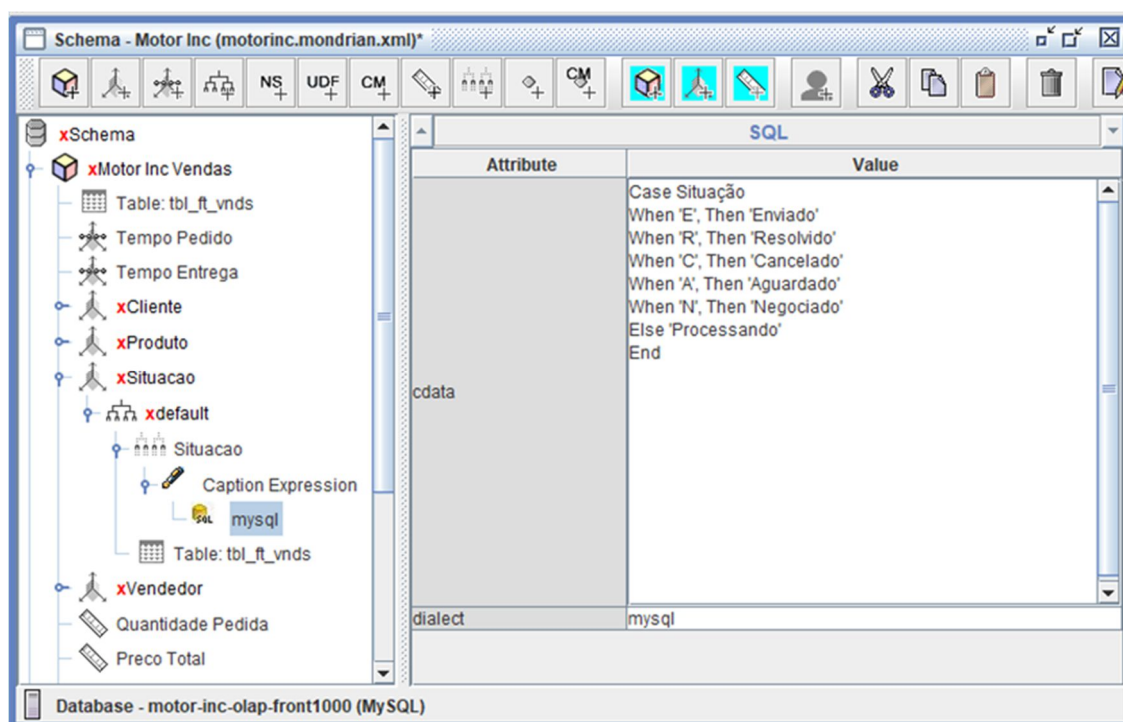


Figura 27-Configuração do atributo Situação- Definição do Caption Expression

6.8.5 Alerta de Medidas

O objetivo desta funcionalidade é mostrar opções de alerta de acordo com as regras de negócio desejada, para este caso foi criado um alerta em cores baseado na medida do Percentual do Valor Total de Vendas. A regra se aplica da seguinte forma, se o percentual de vendas for acima dos 8% então células verdes, se as vendas for abaixo dos 8% então células vermelhas. Este exemplo cria um alerta para determinadas situações tornando-se a informação ao utilizador final mais dinâmica. O resultado pode ser visto na Figura 40. E a configuração na figura 28, logo a seguir.

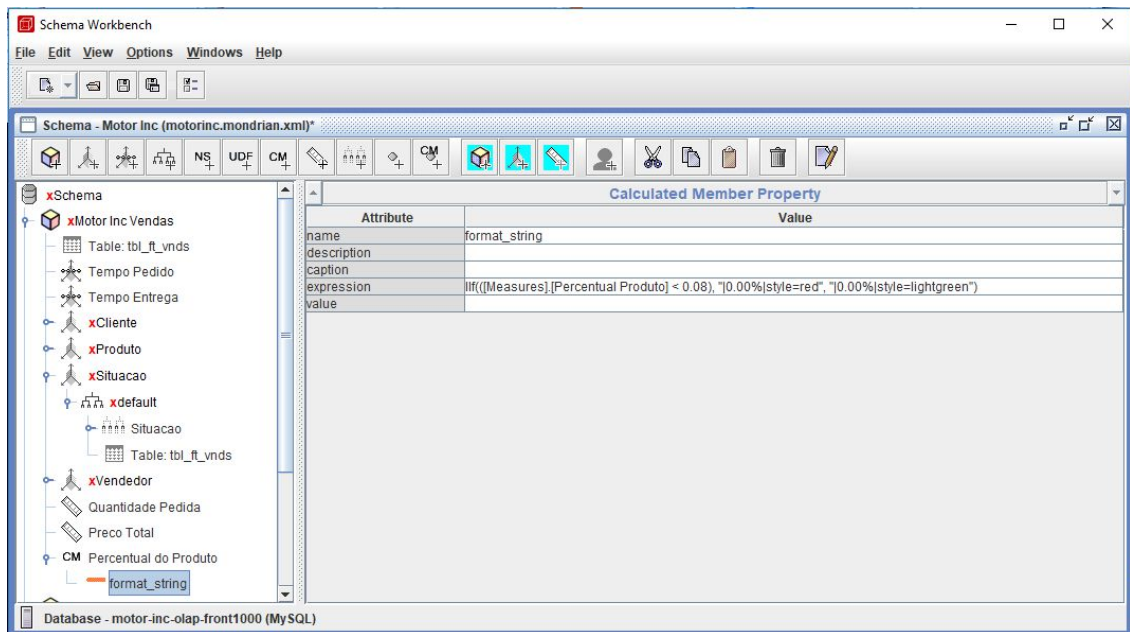


Figura 28- Configuração de alertas para medida- Percentual do valor Total de Vendas.

Clique com o botão direito do mouse na medida Percentual Produto e selecione Add Calculated Member Property e parametrize conforme imagens acima.

Após a configuração do cubo no Schema Workbench pode-se publicá-los no Pentaho BI server para que seja explorado pelos demais utilizadores. Por já existir a conexão com a BD, a publicação dos cubos pode ser realizada através do Mondrian e ser acessado no Pentaho Server. O endereço da web do Pentaho BI Server deve ser especificado na URL. As credencias de qualquer utilizador já cadastrado no Server devem ser as mesmas que as configuradas no PUC.

A figura abaixo demonstra como e feita a publicação do Schema Workbench para o PUC, partindo do menu File > Publish.

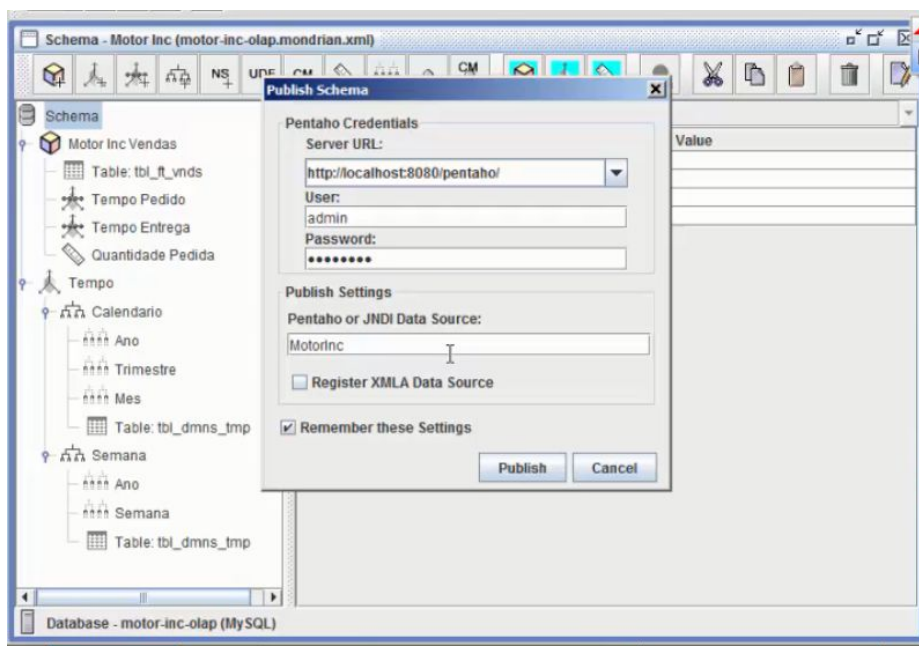


Figura 29-Publicação de um cubo feito no Schema Workbench no BI server

6.8.6 Restrição de Acesso a dados

O Pentaho possui uma funcionalidade é muito importante para qualquer aplicação de BI, que permite configurações na restrição de acesso a documentos e a dados. Neste projeto limitamos a mostrar a restrição de acesso a dados por achar mais relevante. Conceitualmente chamado de Row Level Security, veremos como se aplica esta restrição de acesso a dados em um cubo OLAP.

Primeiramente tem-se que acessar o arquivo `Pentaho.objects.spring.xml` na pasta `biserver/Pentaho-solutions/systems` e fazer a seguinte configuração na linha 264.

```
<bean id="Mondrian-UserRoleMapper"
name="Mondrian-One-To-One-UserRoleMapper"
class="org.pentaho.platform.plugin.action.mondrian.mapper.MondrianOneToOneUserRoleListMapper"
scope="singleton" />
```

Este bean é responsável por verificar qual o utilizador está conectado e qual o seu papel para poder aplicar as restrições. Pode-se aplicar o acesso a nível de cubos também.

A seguir a figura 30 mostra quais são as roles que estão disponíveis no BI Server. O Latam, administrator, Business Analyst, Power User e Report Author. Este grupo de utilizadores não estão disponíveis no Schema Workbench, então temos que referenciá-los para fazer a restrição de acesso aos dados. No caso vamos referenciar os dados ao grupo LATAM e quem for os utilizadores deste grupo vão acessar somente dados do território Latam. No item 6, Criando utilizadores e grupos de utilizadores no BI Server, do Apendice 1 e mostra como é feita a

configuração de permissão de acesso para os utilizadores no Pentaho BI Server que deve estar definido para que o utilizador final possa ser restringido aos dados configurados.

A figura 30 mostra a configuração do utilizador Antonio no grupo LATAM e as operações permitidas.

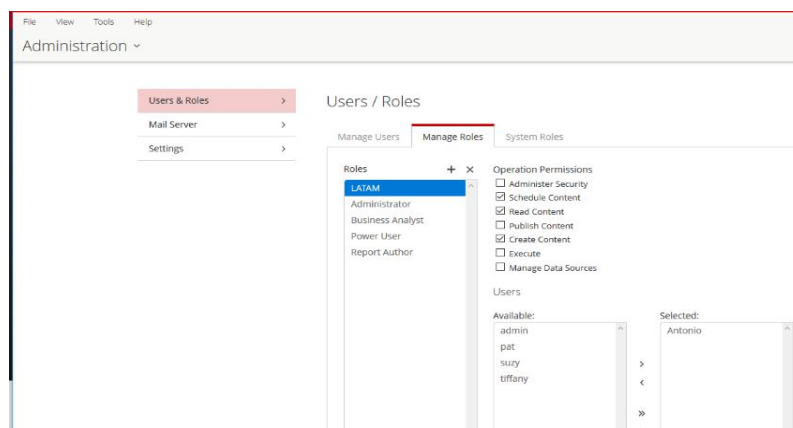


Figura 30-Configuração do utilizador Antonio no Grupo Latam no PUC

Para os variados papéis define-se as seguintes Regras de acesso:

Admin: Acesso a tudo

Authentication: Acesso a nada

Grupo LATAM: Restrição de acesso aos dados na hierarquia Cliente no nível de Território acessado somente os territórios do grupo LATAM.

6.8.6.1 Configuração do Role LATAM no Schema Workbench

Para a configuração do 'Role' LATAM no Schema Workbench, deve-se inserir o Role no mesmo nível de hierarquia da dimensão Tempo pois o utilizador precisará ter acesso a esta dimensão, por isso tem-se que estar no mesmo nível. Dentro da hierarquia LATAM, é inserido o cubo Motor Inc Vendas, do qual o grupo vai ser restringido o acesso. O nível de estrutura hierárquica em que se encontra os dados, deve ser configurado da mesma maneira para que as regras de restrições sejam cumpridas pela ferramenta. O campo a ser aplicado a restrição é o Território, que fica dentro da dimensão Cliente. Podemos ver a configuração da expressão em MDX [Cliente].[LATAM] que será traduzido pelo Mondrian como se fosse um Where de uma Query. A configuração pode-se ver na figura 31 a seguir.

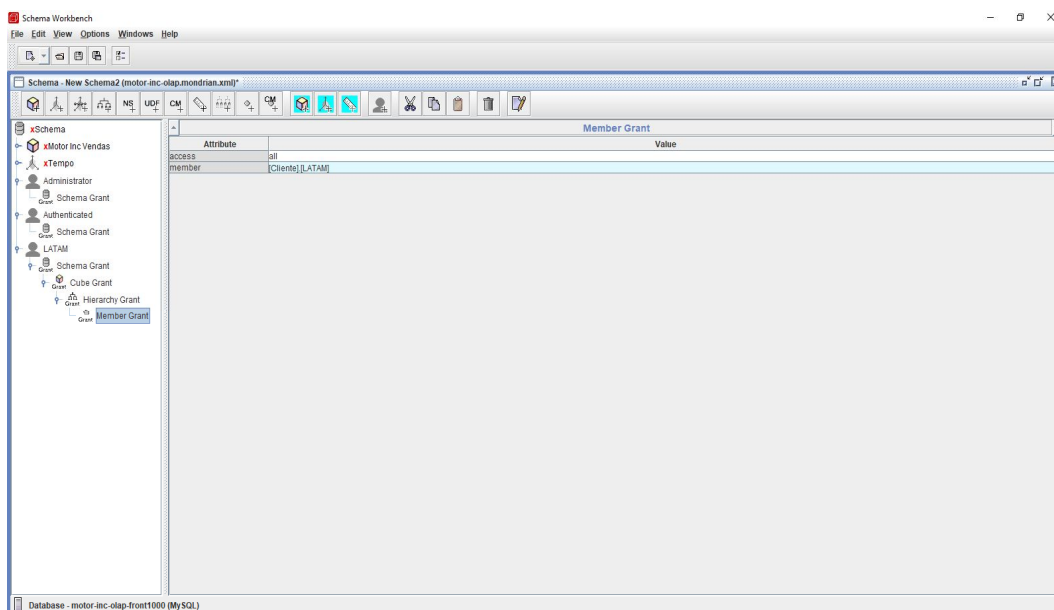


Figura 31-Configuração de restrição de acesso aos dados para o grupo LATAM.

Após essas configurações, deve-se reiniciar o Pentaho BI server e acessá-lo com as credenciais do utilizador Antonio que foi configurado para o grupo Latam, abrir o Saiku e fazer a consulta, quantidade pedida por cliente e território. O resultado da pesquisa trará a restrição para o Grupo LATAM que terá somente dados do território LATAM. Verificar na figura 32.

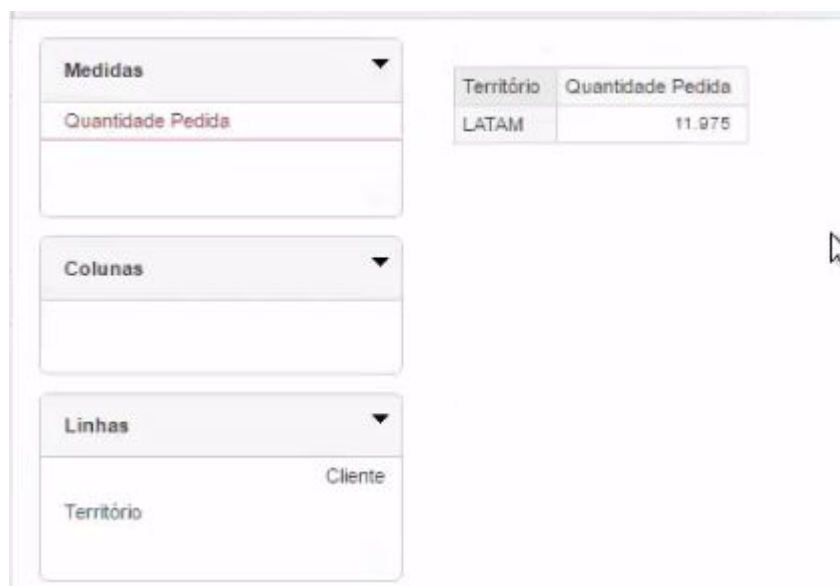


Figura 32-Consulta no Saiku com Restrição de Acesso ao Grupo LATAM.

6.9 BI Server Pentaho

A Interface gráfica do utilizador /A camada de apresentação é representada pelo Pentaho BI Server que disponibiliza artefactos para criação de conteúdos de BI como Relatórios, Dashboards ou consultas sobre cubos OLAP. As funcionalidades oferecidas pela plataforma Pentaho não se limitam somente às aplicações de BI. Esta ferramenta oferece também uma infraestrutura básica de Software e tarefas como autenticação e autorização do utilizador, gerenciamento do conjunto de conexões do banco de dados e execução de tarefas agendadas fazem parte das opções que a ferramenta disponibiliza (Bouman & Dongen, 2009).

O exemplo dos cubos Olap feito anteriormente são exemplos básico de análises que podem ser executadas no Pentaho BI Server. A solução permite ao utilizador a filtragem dos dados e é disponibilizada através dos gráficos e tabelas. Os Relatórios e Dashboards são outros meios de visualizar as informações evidenciando assim a complexidade e o potencial desta ferramenta sendo capaz de produzir diferentes abordagens, inter-relações e interpretações dos dados de uma mesma fonte.

6.9.1 Saiku Analytics

O Saiku Analytics é um cliente web do Pentaho BI Server. Esta ferramenta permite ao utilizador a exploração dos dados de uma maneira rápida bastante simples de arrastar e soltar e uma terminologia de negócio fácil de entender(Esteves, 2016). A proposta é mostrar as análises através dos cubos possibilitando as consultas em visões multidimensionais e a criação de gráficos e tabelas.

Neste exemplo, para efeito de teste da ferramenta, limitamos a criar um cubo com números reduzidos de seleções geradas pelos parâmetros da tabela de vendas Motor inc. Este cubo apresenta o formato da consulta ad-hoc, sendo esta adaptada para agora inserir os resultados numa tabela (cubo), e executada para diversas seleções que pretendemos disponibilizar.

Os resultados das Análises do cubo são apresentados de acordo com as consultas definidas e serão visualizados em dois tipos de perspectivas.Tabela e Gráfico.

Todas as consultas geradas através do cubo no Saiku são criadas na linguagem MDX. Caso o utilizador queira fazer a consulta sem arrastar e soltar pode editar o modo MDX da Saiku como se fosse uma query em uma BD. A linguagem não é muito conhecida, porém na internet pode-se encontrar tutoriais em abundância que explicam como escrever e entender melhor esta linguagem criada pela Microsoft.

A primeira análise, ou consulta exibida na Figura 33, possui as coordenadas linhas e colunas que serão representadas na tabela por Ano e linha de Produto. A visão completa desta medida se define como: Quantidade Vendidas e o Preço Total de Linhas de Produtos por Ano.

6.9.1.1 Quantidade Vendidas e o Preço Total de Linhas de Produtos por Ano.

Na figura abaixo podemos visualizar as 3 zonas disponíveis da Interface como Medidas, Colunas e Linhas. onde foi definida a consulta das vendas de linhas de produtos por ano. Como podemos verificar cada nível é colocado em um eixo que se definiu para Linhas: Linha(s) de produtos, Colunas: Ano em que foi vendido, e Medidas: Quantidade e Preço Total.

Medidas	
Quantidade	
Preço Total	
Colunas	
Ano	Tempo
Linhas	
Linha	Produto

Ano	2003		2004		2005	
Linha	Quantidade	Preço Total	Quantidade	Preço Total	Quantidade	Preço Total
AVIÕES	3.833	347.924	5.820	529.129	2.207	200.077
CAMINHÕES E ÔNIBUS	4.056	420.523	5.024	532.024	1.921	201.903
CARROS CLÁSSICOS	12.762	1.513.998	16.085	1.837.904	6.705	738.587
CARROS VINTAGE	7.913	679.929	20.864	997.708	4.116	388.817
EMBARCAÇÕES	2.844	244.652	4.309	375.498	1.346	128.219
MOTOCICLETAS	4.031	397.392	5.906	590.632	2.771	286.327
TRENS	1.000	72.857	1.409	124.885	409	36.920

Figura 33-Quantidade Vendidas e o Preço Total de Linhas de produtos por Ano- Tabela

Podemos verificar no resultado do filtro que a maior quantidade de produtos vendidos foi no ano de 2003 de carros clássicos, com 12.762 vendas e seu valor total de 1.513.998. É fácil e muito intuitivo para os utilizadores selecionar e analisar o resultado na Interface gráfica do Saiku. No canto superior direito desta Interface encontra-se os dois modos de visibilidade que a tecnologia oferece, a Tabela e o Gráfico.

Existe uma vasta gama de tipos de gráficos disponíveis pela Saiku. Nas duas figuras seguintes apresentaremos o resultado da consulta acima em duas visões diferentes de gráficos. O de pizza e o de barra onde pode ser feito a leitura dos dados com ótima definição.

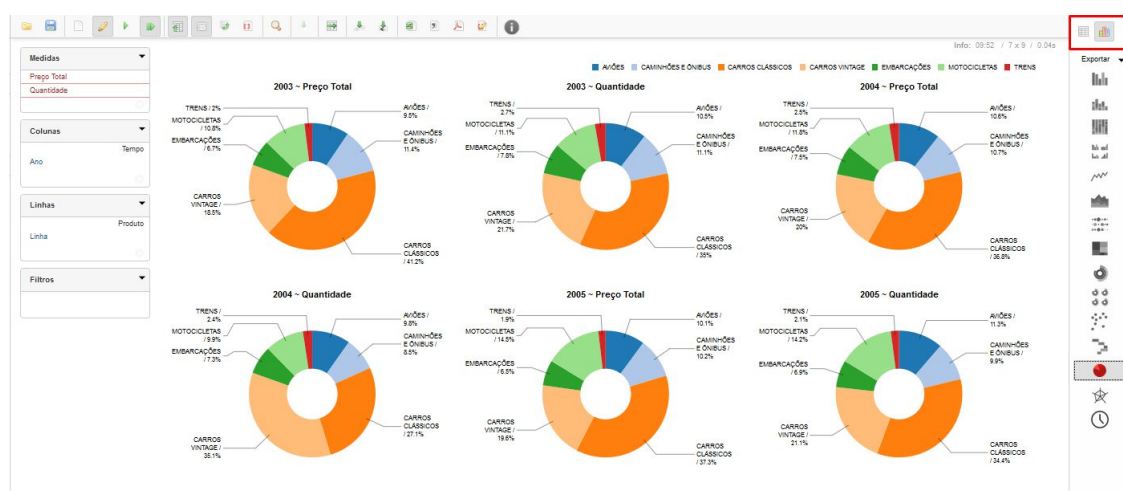


Figura 34-Qtd. Vendas e o Preço Total de Linhas de produtos por Ano – Gráfico de Pizza

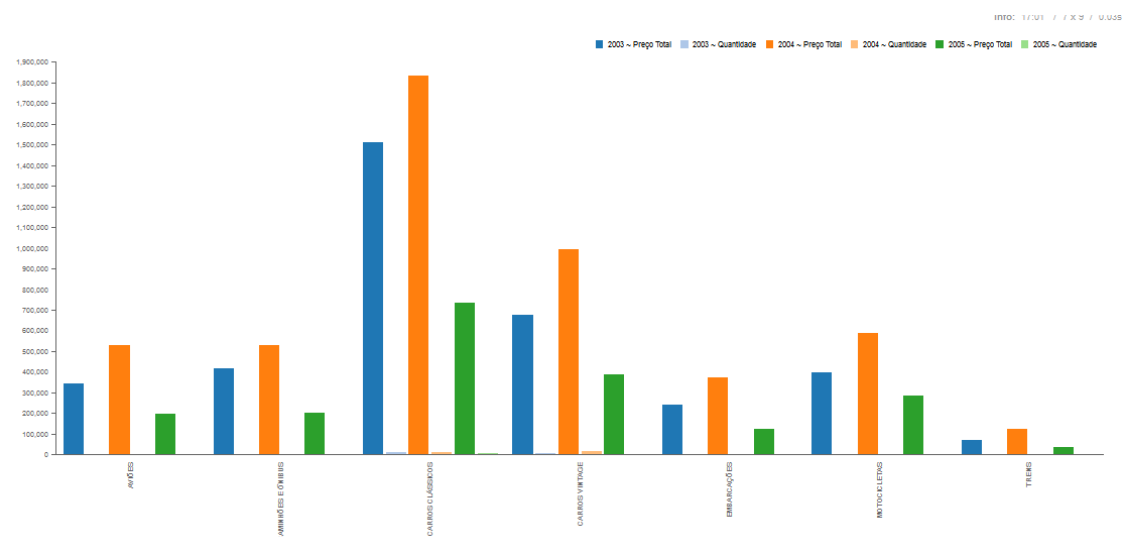


Figura 35-Qtd. Vendas e o Preço Total de Linhas de produtos por Ano – Gráfico de Barra

6.9.1.2 Quantidade e Preço Total de Vendas de Linhas de Produtos por Ano, Situação e País.

Na tabela abaixo pode-se verificar as linhas da tabela representada por Ano, País e Situação, com a categoria Ano representado pelo nível mais alto da agregação. Na tabela abaixo, pode-se ver que os países estão divididos entre os anos 2003, 2004 e 2005. Formando o segundo nível de agregação segue a sequência de países para cada ano. No exemplo do ano de 2005, na categoria países, os dados são novamente subdivididos, desta vez mostrando quase todos atributos do terceiro nível, a Situação: aguardando, em processo, enviado e resolvidos. Por fim, todos os níveis de agregação são divididos por linha de produto.

Esta apresentação é um método claro de níveis de hierarquias que podem ser navegáveis entre os dados, movimentando-se entre os níveis hierárquicos das dimensões, aumentando ou diminuindo seu detalhamento através das operações como Drill down e Roll Up, como pode-se ver. Também pode-se modificar a posição dos dados variando as dimensões ou alternando linha por colunas fazendo o cubo girar com a funcionalidade Pivot ou Rotation.

Medidas																				Info
Preço Total		Ano	País	Linha	AVIÕES		CAMINHÕES E ÔNIBUS		CARROS CLÁSSICOS		CARROS VINTAGE		EMBARCAÇÕES		MOTOCICLETAS		TRENS			
Quantidade	Preço Total				Quantidade	Preço Total	Quantidade	Preço Total	Quantidade	Preço Total	Quantidade	Preço Total	Quantidade	Preço Total	Quantidade	Preço Total	Quantidade	Preço Total		Quantidade
		2003	EUA	Enviado	90.983	977	109.247	1.026	803.284	4.206	289.074	3.091	68.127	702	176.145	1.744	26.876	37		
			AUSTRÁLIA	Enviado	22.316	235	11.301	91	85.355	793	90.133	961	-	-	42.337	431	1.683	3		
			HONG KONG	Enviado	38.617	462	-	-	-	-	5.299	99	-	-	3.850	35	-	-		
			SINGAPURA	Enviado	-	-	44.515	415	79.033	507	15.669	182	14.182	174	-	-	13.273	17		
			ALEMANHA	Enviado	-	-	-	-	55.476	525	11.424	142	-	-	3.139	43	-	-		
			ÁUSTRIA	Enviado	14.221	166	-	-	19.921	188	10.425	117	9.028	113	-	-	-	-		
				Resolvido	-	-	20.860	203	6.726	60	1.350	25	-	-	-	-	-	-		
			BÉLGICA	Enviado	-	-	-	-	-	-	1.640	20	-	-	-	-	1.701	2		
			DINAMARCA	Enviado	-	-	-	-	60.796	478	13.673	157	20.444	230	-	-	4.332	3		
			ESPAÑA	Enviado	21.587	264	67.448	661	129.713	1.072	79.734	921	73.599	835	13.647	193	19.638	29		
			FINLÂNDIA	Enviado	23.111	246	40.469	394	33.983	290	5.945	100	6.309	76	-	-	-	-		
			FRANÇA	Enviado	19.075	219	46.067	467	141.215	1.210	10.731	137	6.660	75	89.175	964	-	-		
			ITALIA	Enviado	54.817	614	3.105	23	43.248	304	47.001	890	6.583	71	4.046	34	3.840	6		
			NORUEGA	Enviado	3.944	29	23.038	232	121.038	1.001	16.372	199	-	-	31.973	294	-	-		
			REINO UNIDO	Enviado	6.990	82	28.127	291	66.636	662	51.149	548	22.415	281	-	-	-	-		
			SUÉCIA	Cancelado	6.912	104	-	-	5.940	44	9.656	115	24.226	287	-	-	-	-		
				Enviado	-	-	-	-	7.221	29	2.630	23	-	-	-	-	-	-		
			FILIPINAS	Enviado	20.926	215	-	-	42.572	391	1.944	27	-	-	12.630	170	-	-		
			BRASIL	Cancelado	17.891	185	-	-	-	-	3.190	55	-	-	5.400	45	-	-		
				Enviado	2.454	36	-	-	29.675	289	15.365	224	-	-	13.050	178	-	-		
			CANADÁ	Enviado	-	-	10.162	96	28.996	225	18.536	177	-	-	-	-	-	-		
			ESTADOS UNIDOS	Enviado	-	-	16.594	165	55.270	526	-	-	-	-	-	-	1.815	3		
		2004	EUA	Cancelado	-	-	-	-	2.920	20	11.766	127	25.570	285	-	-	5.082	4		
				Enviado	102.339	1.130	230.283	2.310	557.490	4.850	255.587	2.734	69.893	802	237.986	2.330	20.513	29		
			AUSTRÁLIA	Enviado	41.416	427	36.171	325	76.198	644	44.325	485	1.080	24	33.077	324	-	-		
			SINGAPURA	Enviado	-	-	44.454	473	54.735	536	16.727	188	-	-	-	-	-	-		
			ALEMANHA	Enviado	23.031	245	10.182	81	92.725	756	9.489	134	5.484	55	4.362	76	5.042	9		
			ÁUSTRIA	Enviado	3.638	34	-	-	15.333	190	6.672	80	-	-	26.060	197	-	-		
			BÉLGICA	Enviado	5.617	61	-	-	3.910	30	31.942	315	31.709	343	-	-	7.315	7		
			DINAMARCA	Enviado	-	-	9.581	73	70.338	592	7.512	83	1.748	23	-	-	7.150	9		
			ESPAÑA	Cancelado	26.497	311	-	-	-	-	16.519	218	6.824	76	-	-	-	-		
				Enviado	14.999	154	54.123	505	209.454	2.070	96.187	1.000	20.601	296	21.294	260	14.726	1		
			FINLÂNDIA	Enviado	-	-	-	-	54.285	389	-	-	-	-	32.177	301	5.136	-		
			FRANÇA	Enviado	29.755	321	40.077	307	194.359	1.704	145.002	1.575	38.747	455	95.359	1.080	16.394	1		
			IRLÂNDIA	Enviado	11.785	115	3.995	37	31.655	202	2.240	29	-	-	4.972	59	3.130	-		
			ITALIA	Enviado	25.063	342	5.492	46	89.937	675	62.003	675	11.089	123	-	-	2.420	-		
			NORUEGA	Enviado	25.569	296	14.060	76	13.059	107	26.650	296	-	-	19.773	190	11.316	-		
			REINO UNIDO	Cancelado	-	-	-	-	50.403	429	-	-	-	-	-	-	-	-		
				Enviado	32.197	397	-	-	42.235	416	72.675	615	47.436	550	-	-	12.628	1		
			SUÉCIA	Aguardando	-	-	20.176	171	6.078	46	-	-	-	-	-	-	-	-		
				Enviado	-	-	27.722	262	33.866	290	21.670	247	6.686	90	-	-	3.908	-		
			SUÍÇA	Enviado	-	-	-	-	117.745	1.075	-	-	-	-	-	-	-	-		
			FILIPINAS	Enviado	-	-	-	-	10.503	87	-	-	-	-	5.432	71	-	-		
			JAPÃO	Enviado	49.202	547	13.356	102	29.412	192	21.482	10.229	10.421	127	26.564	309	-	-		
			BRASIL	Cancelado	-	-	-	-	29.122	196	4.192	32	2.835	33	-	-	5.296	-		
				Enviado	26.258	296	-	-	42.311	467	61.996	682	31.397	339	30.037	216	2.939	-		
			CANADÁ	Enviado	25.529	317	22.351	253	20.600	147	22.826	264	40.303	496	4.182	41	-	-		
			ESTADOS UNIDOS	Enviado	75.164	777	-	-	-	-	31.693	451	7.718	89	49.367	435	-	-		
			EUA	Aguardando	-	-	-	-	2.774	19	16.075	163	23.866	271	-	-	5.908	-		
				em Processo	-	-	-	-	-	-	13.461	151	-	-	-	-	-	-		
				Enviado	17.224	96	41.894	431	149.367	1.216	70.972	650	15.958	190	50.055	519	9.624	1		
				Resolvido	-	-	-	-	19.060	234	25.194	235	-	-	-	-	-	-		
			AUSTRÁLIA	em Processo	-	-	-	-	23.448	288	20.088	244	-	-	-	-	-	-		
				Enviado	7.222	88	29.996	296	9.012	123	27.124	186	-	-	14.478	121	-	-		
				Negociado	2.864	63	-	-	-	-	7.482	69	3.072	32	-	-	-	-		
			SINGAPURA	Enviado	-	-	-	-	-	-	2.616	67	-	-	4.180	44	-	-		
			ÁUSTRIA	Enviado	-	-	-	-	59.450	509	5.775	102	-	-	-	-	-	-		
			BÉLGICA	em Processo	-	-	-	-	-	-	8.420	112	-	-	-	-	-	-		
				Enviado	-	-	-	-	16.631	117	-	-	-	-	-	-	-	-		
			DINAMARCA	Negociado	-	-	-	-	25.974	174	-	-	-	-	-	-	-	-		
			ESPAÑA	em Processo	-	-	17.853	147	15.726	96	1.960	26	-	-	-	-	-	-		
				Enviado	-	-	35.073	396	121.086	1.142	29.545	340	-	-	7.900	65	6.992	1		

Figura 36-Receita de vendas de Linhas de Produtos por Ano, Situação e País-Tabela.

Abaixo temos a versão gráfica da mostragem da evolução da receita das vendas dos produtos, ao longo dos anos de 2003, 2004 e 2005.

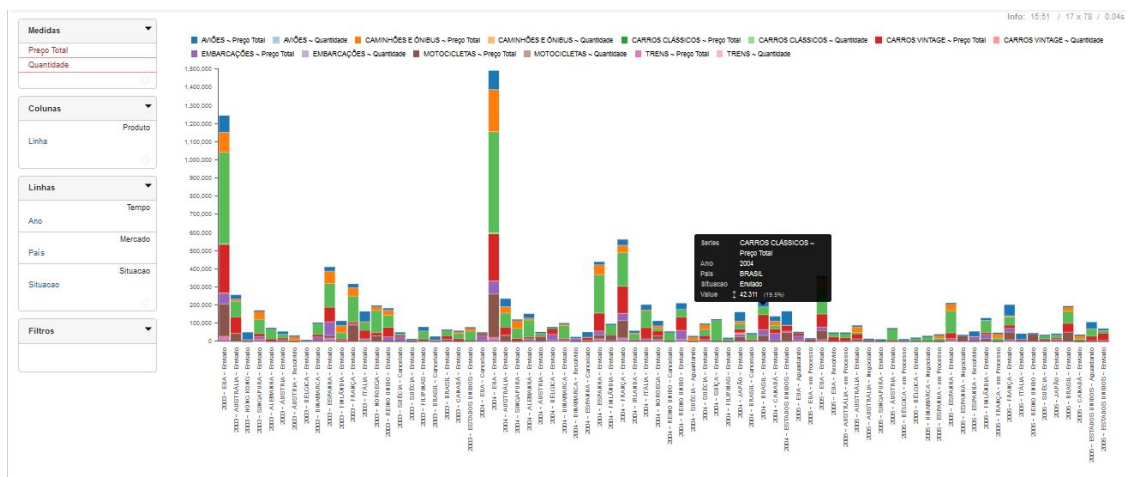


Figura 37-Receita de vendas de Linhas de Produtos por Ano, Situação e País- Gráfico

Slice and Dice é a análise mais básica, porém bem relevante do Pentaho BI Server. O Slice permite selecionar um subconjunto de dados do cubo, através da restrição de uma determinada dimensão. O Dice permite definir um subconjunto através da restrição de mais do que uma dimensão. (Ramos, 2009).

Pivot também conhecida como *rotate* ou rotação, é uma operação de visualização que rotaciona os eixos de um determinado cubo, provendo uma visão alternativa dos dados. A figura 38 mostra uma operação de pivot sobre o resultado da operação de Dice em três dimensões da tabela 36.

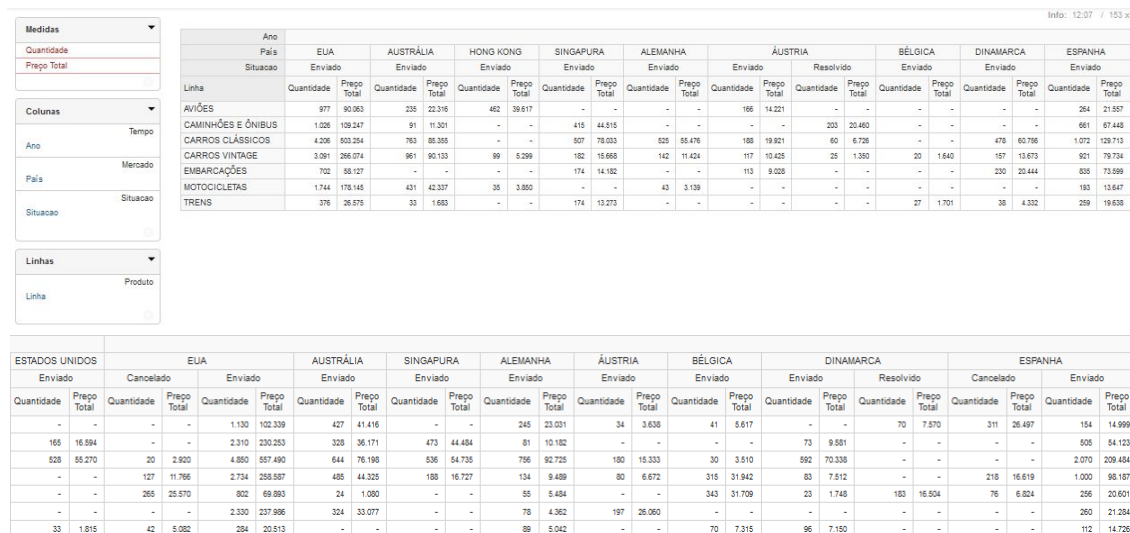


Figura 38- Aplicação Pivot sobre o resultado da operação Dice em três dimensões

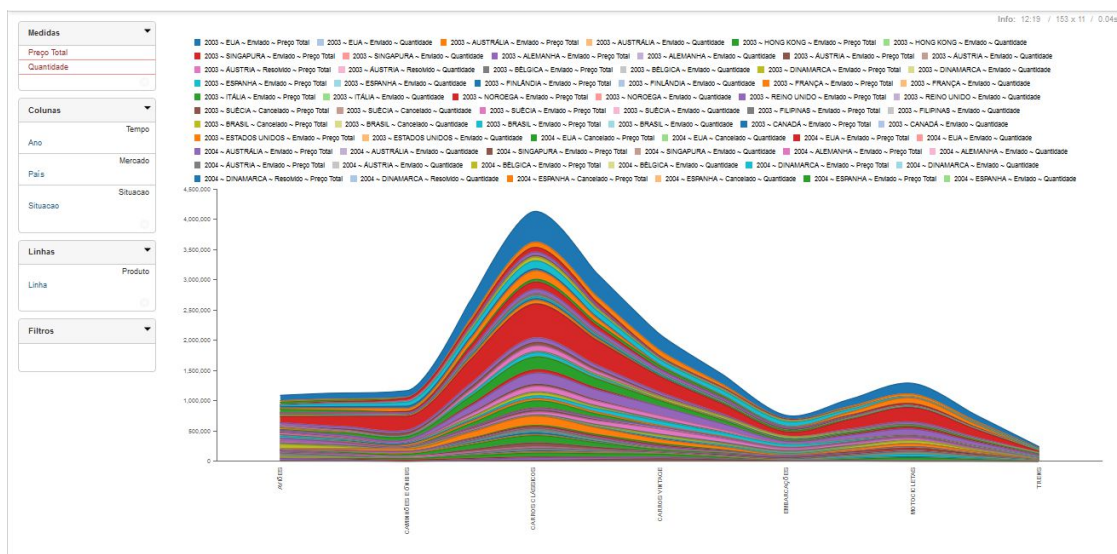


Figura 39- Aplicação Pivot sobre o resultado da operação Dice em três dimensões– Visão Gráfica

6.9.1.3 Quantidade e percentual de Vendas de linhas de Produtos por Fornecedor, por Situação e por Ano.

Como já foi referido antes no item Aplicação de Medidas com destaque(Highlights) no capítulo Schema Workbench, o alerta de cores é aplicado para facilitar a identificação de uma informação que não pode ser implementada pelo front-End. No caso especificado o alerta representa o percentual baseada na medida do Valor Total. A regra se aplica da seguinte forma, se o percentual das vendas for acima dos 8%, aplica-se verde às células correspondentes, se o percentual de vendas for abaixo dos 8% aplica-se vermelho às células correspondentes. Esta funcionalidade é útil pois possibilita o destaque dos valores facilitando a leitura dos dados, principalmente se for aplicado em uma tabela mais complexa.

A figura 40 logo abaixo mostra o resultado das medidas em destaque conforme as regras aplicadas.

Linha	Fornecedor	Situacao		Aguardando				Cancelado				em Processo		Enviado				Negociado	
		Ano		2004		2005		2003		2004		2005		2003		2004		2005	
		Quantidade	Percentual	Quantidade	Percentual	Quantidade	Percentual	Quantidade	Percentual	Quantidade	Percentual	Quantidade	Percentual	Quantidade	Percentual	Quantidade	Percentual	Quantidade	Percentual
AVIÕES	AUTOART STUDIO DESIGN	-	-	-	-	21	100%	24	100%	-	-	-	-	278	100%	406	100%	90	100%
	CLASSIC METAL CREATIONS	-	-	38	100%	47	100%	34	100%	-	-	-	-	318	100%	358	100%	136	100%
	GEARBOX COLLECTIBLES	-	-	42	100%	24	100%	-	-	-	-	-	-	269	100%	425	100%	91	100%
	HIGHWAY 66 MINI CLASSICS	-	-	21	100%	25	100%	-	-	-	-	-	-	255	100%	533	100%	139	100%
	MIN LIN DIECAST	-	-	11	100%	23	100%	40	100%	-	-	-	-	296	100%	371	100%	193	100%
	MOTOR CITY ART CLASSICS	-	-	162	100%	79	100%	81	100%	-	-	-	-	867	100%	1.375	100%	268	100%
	RED eapT DIECAST	-	-	-	-	43	100%	49	100%	-	-	-	-	280	100%	493	100%	119	100%
	SECOND GEAR DIECAST	-	-	92	100%	-	-	83	100%	-	-	-	-	689	100%	953	100%	222	100%
	UNIMAX ART GALLERIES	-	-	52	100%	27	100%	-	-	-	-	-	-	287	100%	457	100%	136	100%
	AUTOART STUDIO DESIGN	42	100%	-	-	-	-	-	-	-	-	46	100%	352	100%	443	100%	143	100%
CAMINHÕES E ÔNIBUS	CAROUSEL DIECAST LEGENDS	34	100%	-	-	-	-	-	-	-	-	18	100%	393	100%	404	100%	148	100%
	CLASSIC METAL CREATIONS	-	-	-	-	-	-	-	-	-	-	11	100%	353	100%	474	100%	107	100%
	EXOTO DESIGNS	-	-	-	-	-	-	-	-	-	-	33	100%	350	100%	505	100%	167	100%
	HIGHWAY 66 MINI CLASSICS	-	-	-	-	-	-	-	-	-	-	18	100%	369	100%	443	100%	149	100%
	MOTOR CITY ART CLASSICS	49	100%	-	-	-	-	-	-	-	-	79	100%	676	100%	839	100%	280	100%
	RED eapT DIECAST	-	-	-	-	-	-	-	-	-	-	44	100%	324	100%	440	100%	131	100%
	STUDIO M ART MODELS	-	-	-	-	-	-	-	-	-	-	54	100%	312	100%	465	100%	131	100%
	UNIMAX ART GALLERIES	40	100%	-	-	-	-	-	-	-	-	38	100%	427	100%	399	100%	143	100%
	WELLY DIECAST PRODUCTIONS	-	-	-	-	-	-	-	-	-	-	49	100%	287	100%	435	100%	148	100%
	AUTOART STUDIO DESIGN	-	-	-	-	-	-	-	-	24	100%	-	-	337	100%	392	100%	190	100%
CARROS CLÁSSICOS	AUTOART STUDIO DESIGN	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Figura 40-Aplicação de Alerta nas células - Tabela

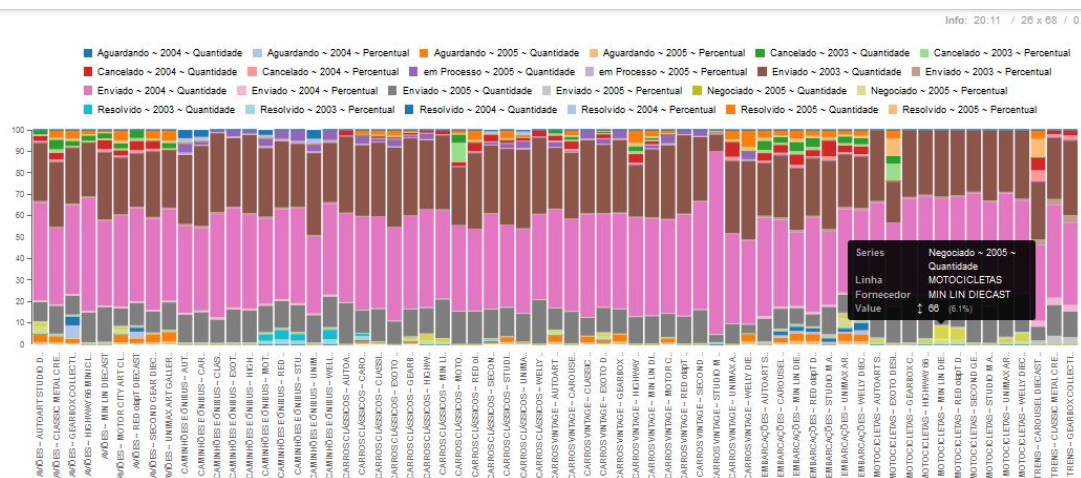


Figura 41-Aplicação de Alerta nas células - Gráfico

O Saiku oferece a possibilidade de exportar os gráficos para formatos como PNG, JPEG e PDF permitindo mais flexibilidade na visualização dos dados ao utilizador final.

6.9.2 Views Dashboards

O CommunityTools Collection, também conhecido como C*Tools é desenvolvida pela Webdetails, uma das mais conceituadas parceiras Pentaho e uma das mais importantes contribuidoras Pentaho do mundo.

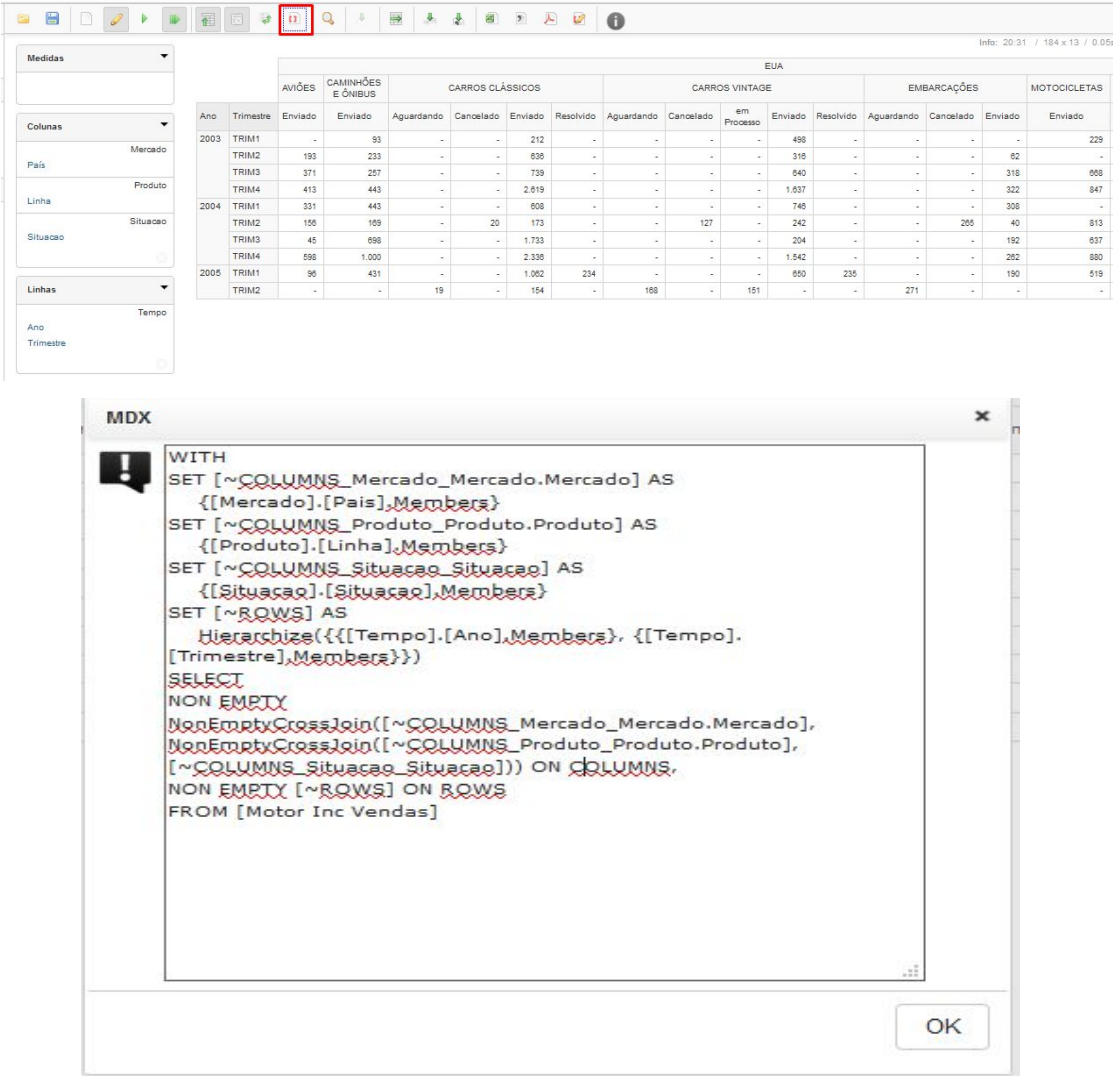
O objetivo das ferramentas C*Tools é melhorar o desenvolvimento de dashboards sólidos e apelativos com Pentaho. (Lima, 2012).

O C*Tools é composta pelos 3 principais componentes CDF(Communtity Dashboards Framework), CDA(Community Data Access), CDE(Community Dashboards Editor).

Nesta sessão limitaremos apenas a mostrar os passos para montar um Dashboard através de uma query MDX e o resultado de alguns Dashboards dos topos de vendas da empresa Motor Inc em uma apresentação mais dinâmica.

6.9.2.1 Criação de um Dashboard através de uma query MDX na Interface CDE

Para exemplificarmos a construção de um Dashboard primeiramente deve-se ir ao Saiku e fazer uma consulta através da montagem de um cubo à escolha, o resultado da query MDX pode ser verificada no ícone identificado em vermelho (MDX) como mostra a Figura 42, logo abaixo.



The screenshot displays the Saiku interface. On the left, there are panels for 'Medidas', 'Colunas', and 'Linhas'. The main area shows a data table with columns for 'Ano', 'Trimestre', and various vehicle categories. A red box highlights the MDX icon in the top toolbar. Below the table, an 'MDX' window is open, showing the following query:

```
WITH
SET [~COLUMNS_Mercado_Mercado.Mercado] AS
{[Mercado].[País].Members}
SET [~COLUMNS_Produto_Produto.Produto] AS
{[Produto].[Linha].Members}
SET [~COLUMNS_Situacao_Situacao] AS
{[Situacao].[Situacao].Members}
SET [~ROWS] AS
Hierarchize({[Tempo].[Ano].Members}, {[Tempo].
[Trimestre].Members}))
SELECT
NON EMPTY
NonEmptyCrossJoin([~COLUMNS_Mercado_Mercado.Mercado],
NonEmptyCrossJoin([~COLUMNS_Produto_Produto.Produto],
[~COLUMNS_Situacao_Situacao])) ON COLUMNS,
NON EMPTY [~ROWS] ON ROWS
FROM [Motor Inc Vendas]
```

The 'MDX' window has an 'OK' button at the bottom right.

Figura 42- Montagem de um Cubo no Saiku e o seu resultado da consulta em MDX

Após copiar a query, já na interface gráfica do CDE e na perspectiva Data Source Panel, no canto superior esquerdo, selecione a opção MDX Queries e preencha as propriedades com os dados abaixo.

Name: MDX Query; Jndi: MotorInc; Mondrian Schema: MotorInc e copie a consulta MDX na opção Query. Todos os passos podem ser vistos com destaque na figura 43.

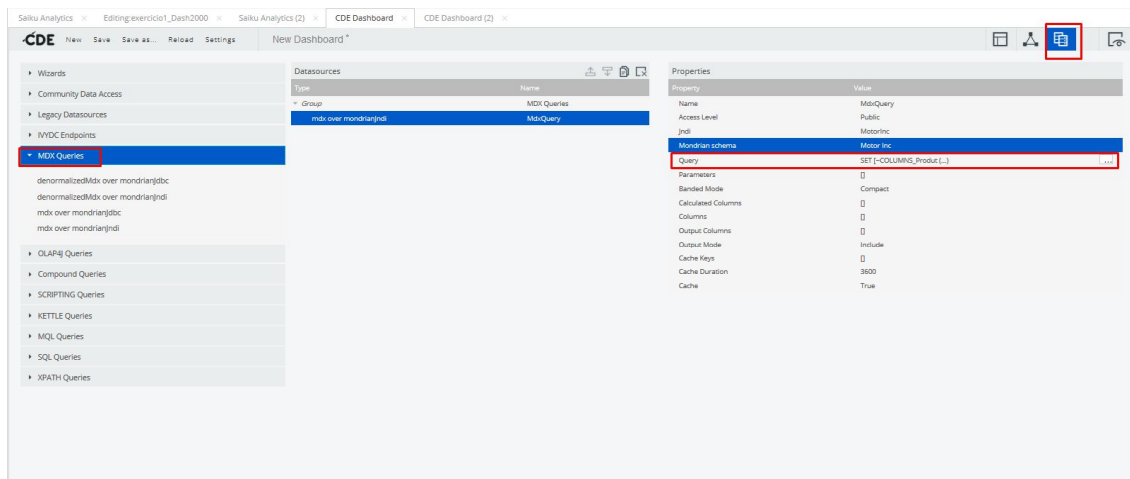


Figura 43-Parametros de entrada para criação do novo Dashboard no CDE

Salve o Dashboard e configure o ficheiro com os parâmetros como mostra a figura 44 a seguir.

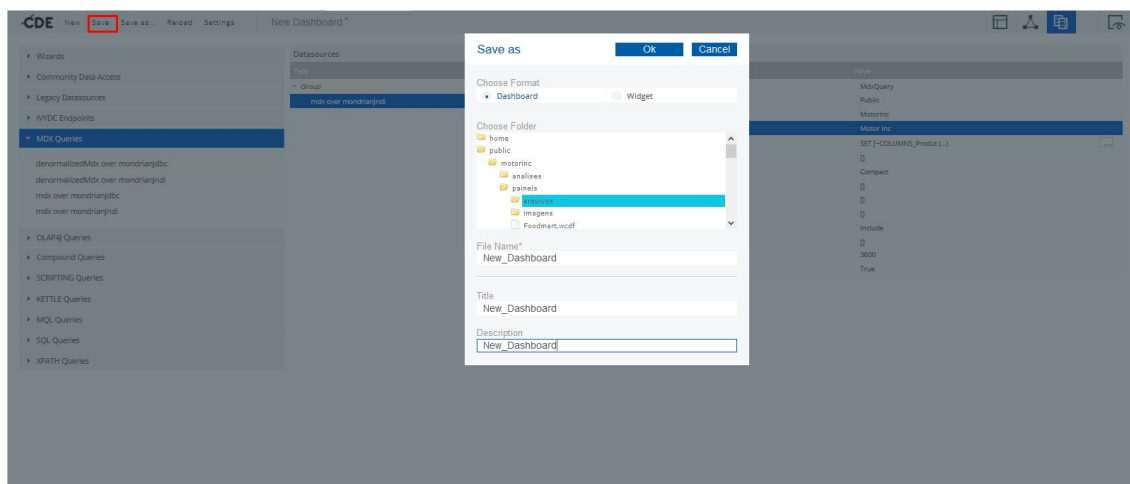


Figura 44- Salvar novo Dashboard CDE

Para finalizar deve-se mudar o tipo do Dashboard CSS, para Blueprint, pois por padrão ele vem no formato Bootstrap e o CDE não reconhece esta versão para a visualização. Deve-se ir ao Menu Setting do CDE e fazer esta configuração como mostra a figura 45 logo abaixo.

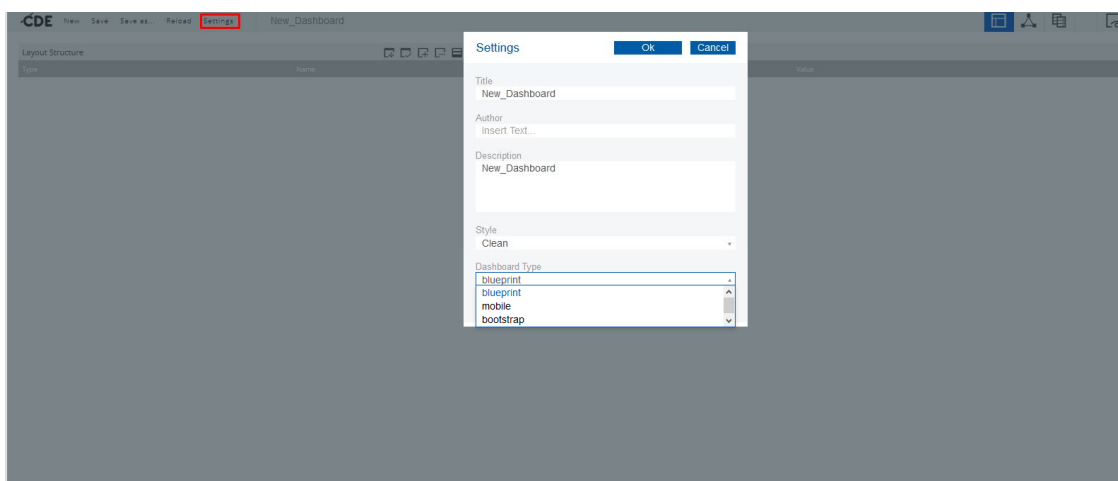


Figura 45-Configuração do Dashboard do tipo Blueprint

6.9.2.2 Resultados dos Dashboards

Abaixo podemos ver o resultado em um painel que mostra 4 Dashboards com os indicadores Top de Vendas por Fornecedor, por Territórios, por Vendedores e por Cidades da empresa Motor Inc.



Figura 46-Performace de vendas por Fornecedores, Territórios e Vendedores

Um outro exemplo de Dashboards mais dinâmico disponibiliza outras formas de visualizações como Relatório à partir do próprio Dashboard. Essas informações podem ser mais detalhadas através de funções Drill down por exemplo. Qualquer um dos níveis selecionados na figura abaixo pode resultar em um relatório.

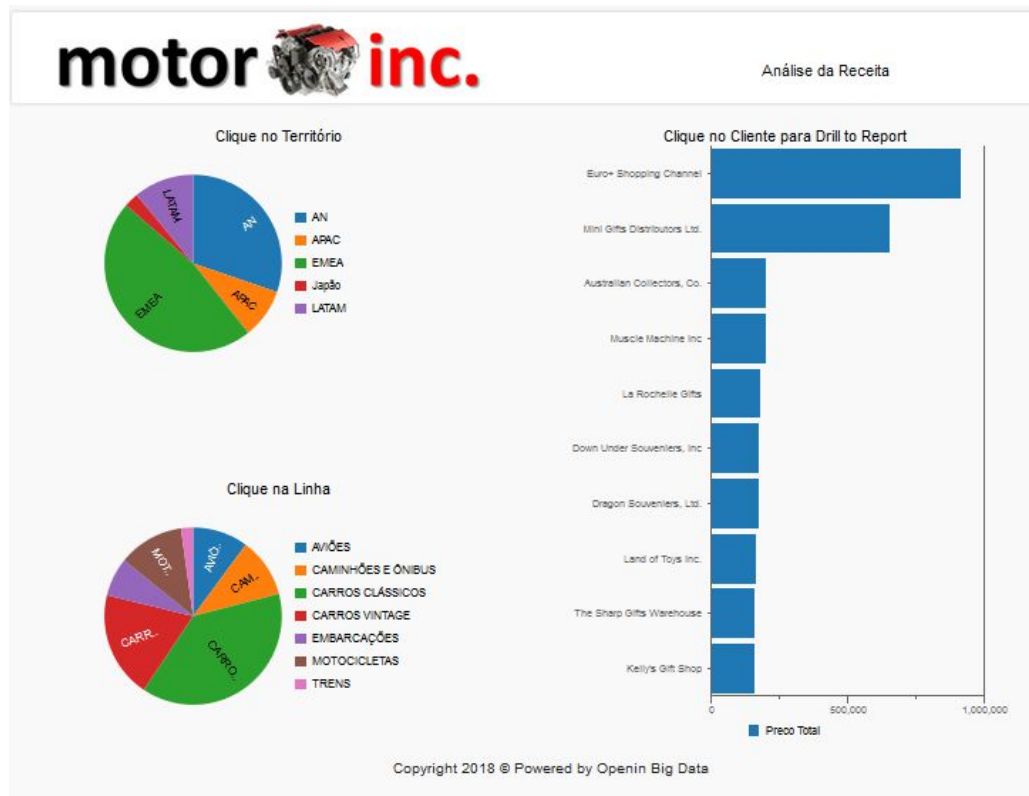


Figura 47-Drillable Dashboard da Receita de vendas da Empresa Motor Inc

Podemos ver na figura 48 a seguir, o resultado das Vendas no Relatório através da seleção do Cliente Mini Gifti Distribution no chart acima.

JAN
Mini Gifts Distributors Ltd.

SKU	Produto	Preço	Quantidade
S12_1108	2001 FERRARI ENZO	3,552.00	32
S12_4473	1957 CHEVY PICKUP	1,764.00	49
S18_2238	1998 CHRYSLER PLYMOUTH PROWLER	2,600.00	25
S24_1444	1970 DODGE CORONET	2,425.00	25
S24_2300	1962 VOLKSWAGEN MICROBUS	3,440.00	20
S24_2766	1949 JAGUAR XK 120	4,120.00	40
S24_2840	1958 CHEVY CORVETTE LIMITED EDITION	5,535.00	45
S24_2887	1952 CITROEN-15CV	4,216.00	31
S24_3191	1969 CHEVROLET CAMARO Z28	3,680.00	46
S24_3432	2002 CHEVY CORVETTE	2,000.00	20
S24_3969	1936 MERCEDES BENZ 500K ROADSTER	1,702.00	46
S24_4048	1992 PORSCHE CAYENNE TURBO SILVER	1,400.00	28
S32_1268	1980'S GM MANHATTAN EXPRESS	4,056.00	26
S32_2509	1954 GREYHOUND SCENICRUISER	1,320.00	20
S32_3207	1950'S CHICAGO SURFACE LINES STREETCAR	3,000.00	30
S50_1392	DIAMOND T620 SEMI-SKIRTED TANKER	2,736.00	48
S700_2824	1982 CAMARO Z28	4,318.00	34
Totais		51,864.00	565

Figura 48- Relatório de Vendas na seleção do Cliente Mini Gifti Distribution.

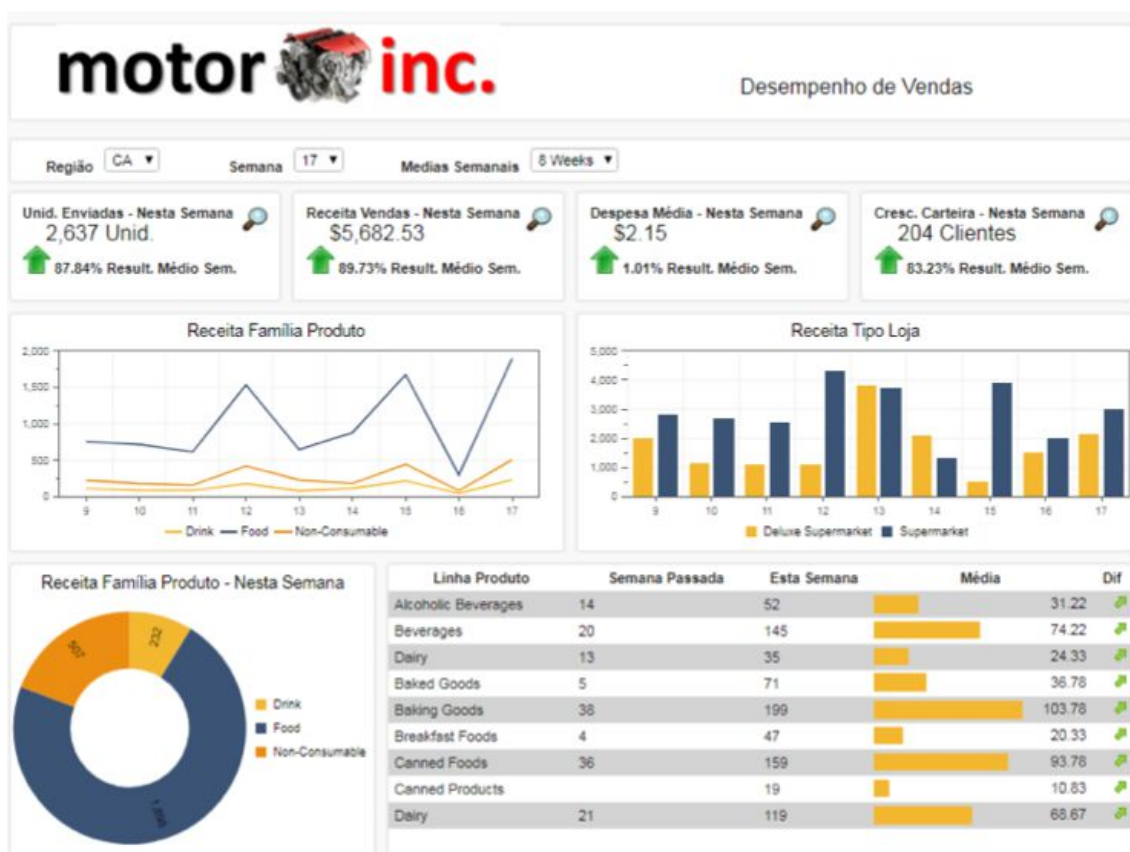


Figura 49- Painel Dinâmico de Dashboards Pentaho-Ctools

6.10 Análise da Suite Pentaho

Com os resultados obtidos através das ferramentas testadas neste projeto, é possível afirmar que se pode hoje em dia obter resultados satisfatórios através de tecnologias de BI Open source. Este tipo de ferramenta possibilita a otimização de Sistemas de BI completo com a entrega de informação correta, consistente, oportuna e acessível. Oferece maior suporte à informação, possibilitando decisões mais rápidas e maior entendimento do impacto destas decisões, diminui claramente os custos e aumenta a receita. Para uma implementação de BI em uma grande empresa, onde exige um processo mais detalhado e mais performático com as ferramentas Open source versão Community, não seria uma boa recomendação pois as ferramentas ainda estão em fase de amadurecimento e ainda precisam de demonstrar mais consistência em algumas operações e performance ao trabalhar com grandes quantidades de dados. Mas existe a possibilidade também de implementação com a versão Enterprise ou até mesmo com a combinação de mais de uma ferramenta no mesmo ambiente, tudo vai depender da necessidade do negócio.

7 Conclusão

O objetivo especificado para este projeto procedeu-se na exploração de algumas das principais ferramentas de Business Intelligence suportadas pela infraestrutura em Nuvem. Com foco nas ferramentas de "Cloud BI" de maior tendência vista por Gartner 2017, foi feito o levantamento das plataformas e análise de suas características gerais, e conseguinte a escolha de uma delas, a Solução Pentaho. Esta fase teve um grande peso na descoberta da quantidade de ferramentas de BI na Nuvem e principalmente com custos reduzidos que se encontra no mercado atual. Qualquer empresa, independentemente da sua dimensão, área de negócio ou posicionamento no mercado, pode beneficiar da implementação de uma plataforma de BI na Nuvem.

Na segunda fase deste projeto foi possível explorar alguns conceitos de Engenharia de Software voltado para a área no domínio acadêmico. Todos os itens relacionados à Análise do negócio só foi possível pois incidiu na avaliação de um documento Excel disponibilizado pela Instituição do Departamento de Informática do Isep e pelo estudo de outros trabalhos existentes relacionados à área no domínio acadêmico.

Todo o restante do projeto teve foco na abordagem dos principais conceitos e processos de BI com as soluções Open Source escolhidas em um cenário no domínio de vendas de produtos automobilísticos.

Depois da fase de concepção e da construção dos modelos dimensionais, procedeu-se a implementação do negócio na área de vendas com foco no principais processos de BI, o ETL e a análise de cubos OLAP, com visualização em tabelas e Gráficos. Utilizando as tecnologias PDI, PSW e o Saiku respectivamente. Importa referir que neste projeto foram as duas fases mais consideráveis devido a importância das funcionalidades na preparação e na análise dos dados.

Devido ao tempo limitado, não foi possível explorar todo o potencial das ferramentas de Visualização tendo somente uma pequena abordagem dos Dashboards e um exemplo de um Relatório.

A escolha das ferramentas mais adequadas ao negócio e a implementação das mesmas depende muito do processo evolutivo na identificação das necessidades que envolve cada fase. Para isso é preciso ter constantemente a comunicação entre os stakeholders, e neste contexto houve uma falha, levando a necessidade da tomada da decisão de dividir o projeto em dois cenários distintos, o que não impossibilitou de atingir o objetivo proposto, de gerir as informações como um todo em um processo centralizado, dispondo de soluções de BI para uma gestão mais eficaz, evitando percas e custos desnecessários.

7.1 Melhorias na Implementação

No que diz respeito ao processo que envolve o sistema de BI, teria sido muito interessante ter implementado o negócio no domínio académico, com as tecnologias que envolve o processo ETL, Olap Analysis, Relatórios e Dashboards. Não foi possível o seguimento na implementação devido à dificuldade encontrada no início, através das análises dos dados disponíveis para a definição do âmbito da aplicação e consequentemente na especificação do modelo. No entanto, ficam aqui expressas as ideias para futuras versões de implementação da plataforma. Também podemos acrescentar para o negócio como um todo a implementação das funcionalidades na área de:

- Análise das prescrições ao longo dos anos letivos, tendo em conta as normas em vigor aplicadas. Com esta análise seria possível obter informação sobre os cursos que tinham mais prescrições, características em comuns dos alunos que prescreviam e em que ano de estudo havia mais tendências a prescrever.
- Com os dados de assiduidade guardados dos alunos é possível fazer um acompanhamento das presenças dos alunos nas disciplinas permitindo a análise e descobertas sobre quais unidades curriculares a que os alunos faltam mais, quais as aulas mais frequentadas, quais as aulas que possuem mais assistência até o fim, quais os alunos que deixam de frequentar as aulas e quais suas causas.
- Satisfação dos alunos: Inquérito simples de satisfação para o aluno de cada disciplina e avaliação do desempenho do professor, na didática, no conhecimento, e nas avaliações das provas, com espaço para observações. Pretendia-se analisar a opinião dos alunos de modo a combater a sua insatisfação e o melhoramento na gestão dos docentes e automaticamente no sucesso dos cursos.

7.2 Dificuldades Encontradas

- Dificuldade em encontrar documentação atualizada(inexistente) mais completas na internet e mesmo nos fóruns específicos sobre as ferramentas BI Pentaho e seus componentes, tornando-se difícil perceber suas configurações, integrações e implementações.
- Dificuldade em perceber a definição de cada atributos para a implementação do modelo dimensional no negócio de gestão acadêmica.
- Dificuldade na implementação do modelo com variadas tecnologias. Muita tecnologia a ser estudada para a implementação do negócio, sem orientação e o tempo limitado ficou árduo.
- Dificuldade na Tomada de decisão na estruturação da documentação. Do que seria útil e de mais valia mostrar na análise do negócio na implementação das tecnologias estudadas, sem ter que sair do contexto de uma estrutura de relatório de uma tese.
- Dificuldade na comunicação com os stakeholders.

8 Referências

Alsufyani, R. & Chang, V., 2015. *RISK ANALYSIS OF BUSINESS INTELLIGENCE IN CLOUD COMPUTING*. [Online].

Amazon, W. S., 2017. *Amazon Elastic Computing Cloud*. [Online]
Available at: <https://aws.amazon.com/pt/ec2/>

Ana Medeiros, A. C. F. V., 2010. *Apoema - Gestão de um portal para estagios de estudantes*.

Analytics, I., 2017. *IBM Cognos software*. [Online]
Available at: <https://www.ibm.com/analytics/us/en/technology/cognos-software/>

Azure, 2017. *Elasticidade na Cloud*. [Online]
Available at: <https://azure.microsoft.com/pt-pt/overview/what-is-elastic-computing/>

Azure, M., 2016. *Microsoft Azure*. [Online]
Available at: <http://www.microsoft.com/azure/>.

Azure, M., 2017. [Online]
Available at: <https://azure.microsoft.com/pt-pt/overview/what-is-iaas/>

Bakshi, A. & Dujodwala, Y. B., 2010. *Securing cloud from ddos attacks using intrusion detection system in virtual machine*.

Birst, F. O., 2017. *Finances Online*. [Online]
Available at: <https://reviews.financesonline.com/p/birst/>

Bouman, R. & Dongen, J. v., 2009. *Pentaho Solutions Business Intelligence and Data Warehousing with Pentaho and MySQL*. s.l.:WILEY.

Bucksch, I. H. S., 2016. *Cloud Computing in SMEs*.

Capuanol, E. A. et al., 2009. *Inteligência competitiva e suas conexões epistemológicas com gestão da informação e do conhecimento*. [Online].

CARVALHO, C., MONTARDO, S. P. & ROSA, H. A., 2008. *Blogs como ferramentas de gerenciamento de Imagens das Organizações*.

Castro, J. M. L. T. d., 2016. *Tendências de Bisuness Intelligence*.

Castro, J. M. L. T. d., 2016. *Tendências de Business Intelligence*.

CENESP, s.d. *Soluções para gestão na Nuvem*. [Online]
Available at: <http://enginebr.com.br/3733.html>

Chang, V. & Alsufyani, R., 2015. *RISK ANALYSIS OF BUSINESS INTELLIGENCE IN CLOUD COMPUTING*.

Chaudhuri, S., Dayal, U. & Narasayya, V., 2011. *An Overview of Business Intelligence Technology*. [Online].

- Chentouf, B., 2007. *Smile Les Blog des Experts*. [Online]
Available at: <http://blog.smile.fr/Pentaho-ou-spagobi>
- Ciurana, E., 2009. *Developing with Google App Engine*.
- Cognos, F. O. I., 2017. *FinancesOnline*. [Online]
Available at: <https://reviews.financesonline.com/p/ibm-cognos/>
- Costa, S. & Santos, M. Y., 2009. *Sistema de Business Intelligence no suporte à Gestão Estratégica*. [Online].
- CRISTESCU, M. P., 2016. *Traditional Enterprise Business Intelligence Software Compared to Software as a Service Business Intelligence*. [Online].
- Devmedia, J. L., 2016. *Conhecendo a Plataforma, Arquitetura e Infraestrutura*. [Online]
Available at: <http://www.devmedia.com.br/pentaho-bi-conhecendo-a-plataforma-arquitetura-e-infraestrutura/31502>
- DOMO, R., 2017. *Domo Review*. [Online]
Available at: <https://reviews.financesonline.com/p/domo/>
- Dongen, R. B. J. v., 2009. *Pentaho Solutions - Business Intelligence and Data Warehousing with Pentaho and MySQL*. s.l.:s.n.
- Dongen, R. B. J. v., 2009. *Pentaho Solutions Business Intelligence and Data Warehousing with Pentaho and MySQL*. s.l.:s.n.
- Eckerson, W. W., 2005. *The Keys to Enterprise Business Intelligence: Critical Success Factors*, The Data Warehousing Institute.
- Edmonds, A., 2010. *Open Cloud Computing Interface*.
- Esteves, A. M. R., 2016. *Plataforma Web para monitorização de dados em contexto escolar*, s.l.: s.n.
- Fengyu Wang, A. Y. L. Y., 2008. *Combined Application of Cloud Computation Technology and Business Intelligence*.
- Frainer, J., 2010. *FERRAMENTAS DE COLABORAÇÃO E GERENCIAMENTO TECNOLÓGICO DA INFORMAÇÃO EM EMPRESAS DE TECNOLOGIA*.
- Frainer, J. & Fontana, G. A., 2010. *Ferramenta de colaboração e gerenciamento tecnológico da Inf. em empresas de TI-*.
- Freitas, A. d. A., 2015. *Pentaho Agregando valor ao seu negócio*. [Online]
Available at: <https://pt.linkedin.com/pulse/pentaho-agregando-valor-ao-seu-neg%C3%B3cio-andressa-de-andrade-freitas>
- Gartner, 2016. *Gartner Magic Quadrant*. [Online]
Available at: <http://www.gartner.com/technology/research/methodologies/magicQuadrants.jsp>
[Acedido em 15 02 2016].

- Gartner, 2016. *Magic Quadrant for Data Integration Tools*. [Online]
Available at: <https://www.gartner.com>
- Gartner, 2017. *Gartner*. [Online]
Available at: <http://www.gartner.com/newsroom/id/3612617>
- Gomes, M. N. d. S., 2013. *Estatística e Gestão da Informação*. [Online]
Available at: <https://run.unl.pt/bitstream/10362/10513/3/TEGI0346.pdf>
- Gonçalves, P. M. F. D., 2012. CLOUD COMPUTING SOBRE A PLATAFORMA.
- GONÇALVES, P. M. F. D., 2012. *Cloud Computing Sobre a Plataforma Windows Azure*. [Online].
- Gonzales, N. et al., 2012. A Quantitative Analysis of Current Security Concerns and Solutions for Cloud Computing.
- Gonzalez, N. M. et al., 2013. Segurança das Nuvens Computacionais.
- GoodData, F. O., 2017. *Finances Online*. [Online]
Available at: <https://reviews.financesonline.com/p/gooddata/>
- Hamburgo, I. & Buksch, S., 2016. *Cloud Computing in SMEs*. [Online].
- Hill, R., 2007. Eucalyptus: a web service-enabled e-infrastructure.
- InQuesti, 2016. *InQuesti*. [Online]
Available at: <http://www.inquesti.com.br>
- Insider, B., 2016. Here's how the Internet of Things will explode by 2020.
- Institute, I. G., 2010. *COBIT Modelo*. s.l.:s.n.
- Jaspersoft, F. O., 2017. *Finances Online*. [Online]
Available at: <https://reviews.financesonline.com/p/jaspersoft/>
- Kimball, R. & Ross, M., 2002. *The Data Warehouse, Toolkit The Complete Guide to Dimensional Modeling*. Second Edition ed. s.l.:John Wiley & Sons, Inc..
- Koen, P. A. e. a., 2001. Providing clarity and a common language to the “fuzzy front end”. Research Technology Management.
- LAPA, J. M. O., 2015. Estudo e Avaliação de Plataformas de Business Intelligence em contexto empresarial.
- LAPA, J. M. O., 2015. Estudo e Avaliação de Plataformas de Business Intelligence em contexto empresarial.
- Lima, M. C. C., 2012. *BUSINESS INTELLIGENCE NO ENSINO SUPERIOR*, Instituto Politecnico de Viana do Castelo : s.n.
- Li, N. et al., 2013. systems, A new methodology to support group decision-making for IoT-based emergency response.

Li, N., Sun, r., BiZeya, Z. & Wang, S., s.d. *A new methodology to support group decision-making for IoT-based emergency response systems.* [Online] [Acedido em 2013].

MACHADO, F. N. R., 2007. *Tecnologia e projeto em Data Warehouse.* São Paulo: Editora Érica.

Menzel, M. & Rajan, R., 2016. *CloudGenius: decision support for web server cloud migration.* s.l., s.n.

Microsoft, 2017. *serviços do Powe rBI.* [Online] Available at: <https://powerbi.microsoft.com/pt-br/documentation/powerbi-frequently-asked-questions/>

Mircea, M., Ghilic-Micu, B. & Stoica, M., 2007. Combining Business Intelligence with Cloud Computing to delivery agility in actual economie.

Moreira, V. H. M., 2015. ViDi Stat.

Nelson M. Gonzalez, C. C. M. F. F. R. M. S., 2011. A Taxonomy Model for Cloud Computing Services.

Online, F. P., 2017. *Finances Online.* [Online] Available at: <https://reviews.financesonline.com/p/microsoft-power-bi/>

OpenI, 2011. [Online] Available at: https://openi.nlm.nih.gov/detailedresult.php?img=PMC3222190_jmir_v13i3e67_fig1&req=4

Pentaho, C. H. G., 2016. *Pentaho Hitachi Group Cmpany.* [Online] Available at: <http://www.pentaho.com/resources/fr>

Popovic, A., Coelho, P. & Jaklic, J., 2009. The impact of business intelligence system maturity on information quality..

Povo, G. d., 2016. *Como a computação na nuvem está tomando conta do mercado.* [Online] Available at: <http://www.gazetadopovo.com.br/economia/como-a-computacao-na-nuvem-esta-tomando-conta-do-mercado-2yeffa82ip57lfuq9o4n1cncz>

PowerBI, F. O., 2017. *Finances Online.* [Online] Available at: <https://reviews.financesonline.com/p/pentaho/>

Ramos, M. Y. S. e. I., 2009. *Business Intelligence.* s.l.:Editora de Informática.

Reguia, C., 2014. Product innovation and the competitive advantage. *European Scientific Journal.*

Reis, T. B. & Silva, P. C. d., 2014. AN ANALYSIS OF THE CLOUD COMPUTING APPROACH TO SOA.

Reis, T. B. & Silva, P. C. d., 2014. AN ANALYSIS OF THE CLOUD COMPUTING APPROACH TO SOA.

REVIEW, B., 2017. *Birst REVIEW Business Intelligence.* [Online] Available at: <https://reviews.financesonline.com/p/birst/>

Rita L. Sallam, C. H. C. J. I. T. W. O., 2017. Magic Quadrant for Business Intelligence. 16 February, p. 30.

Ruan, V. & Baars, H., 2013. Decision support for partially moving applications to the cloud: the example of business intelligence..

Sano, M. D., 2014. *Business Intelligence as a Service: a new approach to manage business processes in the Cloud*. [Online].

SAS, F. O., 2017. *Finances Online*. [Online]
Available at: <https://reviews.financesonline.com/p/sas-business-intelligence/>

Scharf, E. R. & Soriano-Sierra, E. J., 2008. A gestão do conhecimento e o valor percebido.

Sisense, F. O., 2017. *Finances Online*. [Online]
Available at: <https://financesonline.com/top-3-business-intelligence-software/>

Solutions, P. p. O. S., 2014. *Guide Open Source by Smile*. [Online]
Available at: <http://www.open-source-guide.com/en/Solutions/Applications/Business-intelligence-suites/Pentaho>

Sousa, F. R. C., Moreira, L. O. & Machado, J. A. F. d. M. e. J. C., 2011. Gerenciamento de Dados em Nuvem: Conceitos, SistemaseDesafios.

Sousa, F. R. C., Moreira, L. O. & Machado, J. C., 2010. Computação em Nuvem: Conceitos, Tecnologias,.

SpagoBI, 2016. [Online]
Available at: <https://www.spagobi.org>

Standardization., I. O. f., s.d. *ISO/IEC 9126 - Software Product Quality - Part 1: Quality Model*. s.l.:s.n.

Standardization., I. O. f., s.d. *ISO/IEC 14598 - Software Product Evaluation - Part 4: Process for Acquirers*,. s.l.:s.n.

Station, I. C., 2017. [Online]
Available at: https://www.itcentralstation.com/products/comparisons/pentaho_vs_spagobi

Stipié, A. & Bronzin, T., 2012. How Cloud Computing Is (not) Changing the Way We Do BI.

Tableau, 2016. *Tableau*. [Online]
Available at: <https://www.tableau.com>

Tableau, F. O., 2017. *Finances Online*. [Online]
Available at: <https://reviews.financesonline.com/p/tableau-software/>

Takashina, N. T., 1998. *Indicadores de Qualidade e de desempenho*. [Online]
Available at: http://cliente.argo.com.br/~mgos/indicadores_da_qualidade_takashina.htm

team, f., 2017-11-23. <https://reviews.financesonline.com/>. [Online].

Tereso, M., 2011. Ferramentas de BI Open Source para PME's.

US, C., 2017. <https://reviews.financesonline.com/>. [Online].

Van Leemput, E., 2014. Internet of Things (IoT) Business Opportunities – Value.

View, R. Q., 2017. *Qlik View Review*. [Online]
Available at: <https://reviews.financesonline.com/p/qlikview/>

Westcom, 2016. *Westcom Delivering Results Together*. [Online]
Available at: <http://blogbrasil.westcon.com/como-gerar-receita-recorrente-com-a-nuvem>

Yan, L., Rong, C. & Zhao, G., 2009. Cloud Computing Security with Federal Identity Management Using Hierarchical Identity-based Cryptography.

Zuza, E. d. S., 2008. Revista Brasileira de Inovação Científica em Comunicação.

Apêndice A- Configurações

1. Criando e editando parâmetros de conexão

Esta configuração servirá para todas as conexões feitas entre as ferramentas da Suite Pentaho e uma BD (motor-inc-Oltp/Olap).

- Edite o arquivo `jdbc.properties` que está localizado na pasta `...\designtools\data-integration\simple-jndi`;
- Adicione o seguinte bloco de texto abaixo, ao final do texto no arquivo:

```
des_star/type=javax.sql.DataSource
des_star/driver=com.mysql.jdbc.Driver
des_star/url=jdbc:mysql://localhost:3306/motor-inc-star
des_star/user=root des_star/password=root
```

- Salve o ficheiro

2. Configuração do ficheiro properties no Mondrian

O Mondrian quando vai buscar as informações sobre consulta ou na cache ou no Banco, ele não verifica por padrão se existem tabelas agregadas, ele lê somente as tabelas de facto, por isso deve-se habilitar esta ação à mão.

Ir ao arquivo `mondrian.properties` do `pentaho-server/Pentaho-solutions/system/Mondrian`, e configure estas duas linhas:

- `mondrian.rolap.aggregates.Use=true`
- `mondrian.rolap.aggregates.Read=true`
- Salve o arquivo

3. Configuração de log no Mondrian

E para se fazer um melhor acompanhamento das ações do Mondrian em suas consultas, deve-se habilitar os logs que se encontram no ficheiro `log4j.xml` na pasta

Pentaho-server/tomcat/Pentaho/web-inf/classes e configure o código a seguir

```
<appender name="SQLLOG"
class="org.apache.log4j.RollingFileAppender">
  <param name="File" value="../../logs/mondrian_sql.log"/>
  <param name="Append" value="false"/>
  <param name="MaxFileSize" value="500KB"/>
  <param name="MaxBackupIndex" value="1"/>
  <layout class="org.apache.log4j.PatternLayout">
    <param name="ConversionPattern" value="%d %-5p [%c]
%m%n"/>
  </layout>
</appender>
<category name="mondrian.sql">
  <priority value="DEBUG"/>
  <appender-ref ref="SQLLOG"/>
</category>
```

4. Conexão JDBC com a Base de Dados Schema Workbench

Preencha os campos adequadamente com os dados de conexão. Veja a figura.

Menu Option > connection:

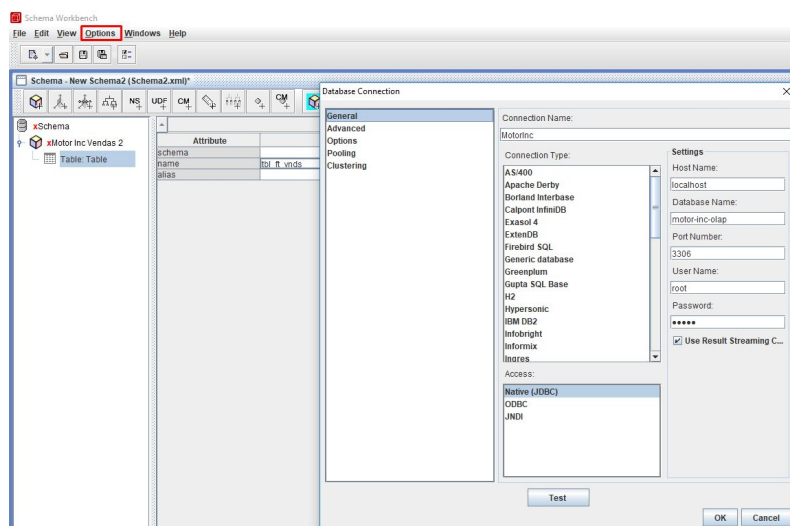


Figura 50-Configuração conexão Base de Dados JDBC/JNDI

PS. Os dados de conexão desta configuração é padrão para todos os componentes utilizados aqui e que integraram/conectaram a uma Base de Dados. PDR, PSW, PDI, BI Server, Saiku, CDA.

5. Abrindo Cubo no Saiku - PUC

Depois de publicar o cubo no Bi server, para abri-lo no Saiku siga as imagens abaixo

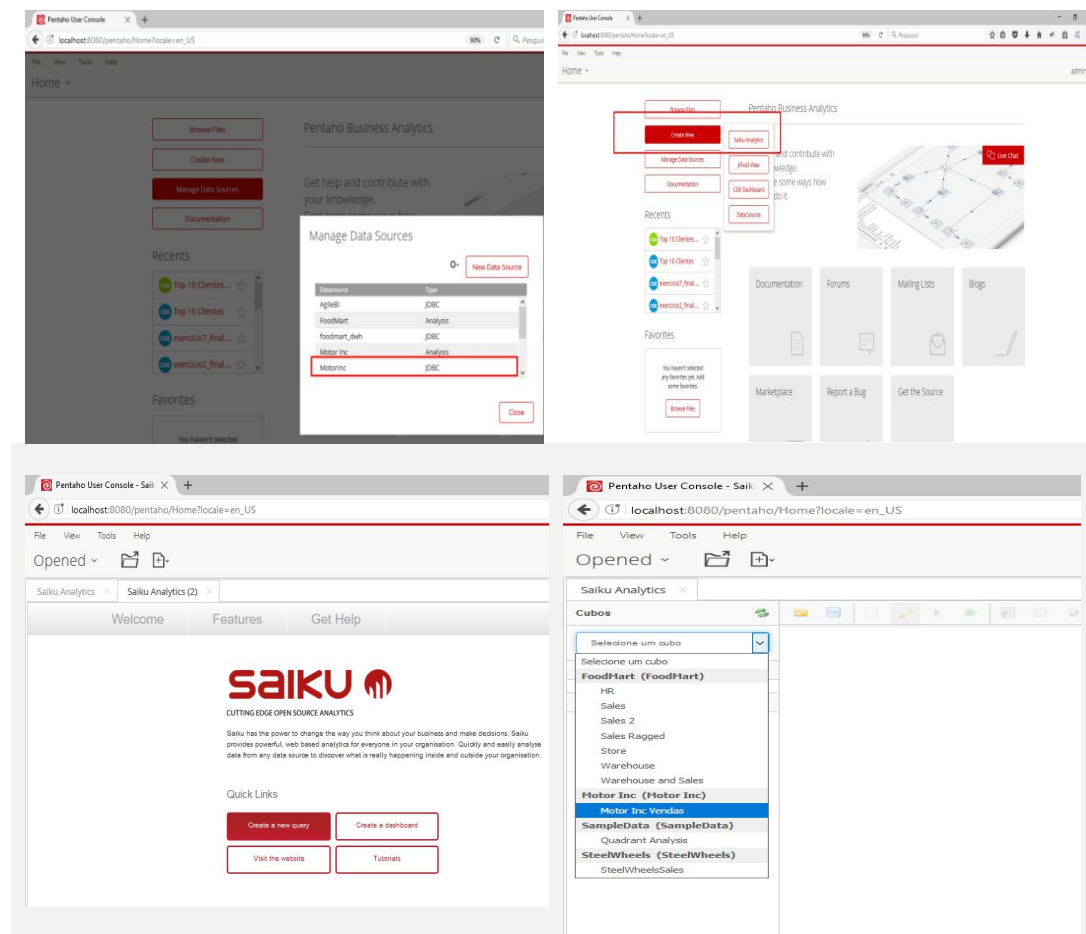


Figura 51-Sequencia ilustrativa de como abrir cubo no Saiku

6. Criando utilizadores e grupos de utilizadores no BI Server

Home > Administration > Users & Roles >

Aba: Manage Users :

Users (Admin)ex.: New User: Antonio; password:123;

Aba: Manager Roles (botão +): LATAM > Permissões:

Escolha das opções: Agendamento, ler conteúdo, publicar conteúdo, criar conteúdo;

P.S. Os conteúdos são salvos automaticamente no JackRabitt.

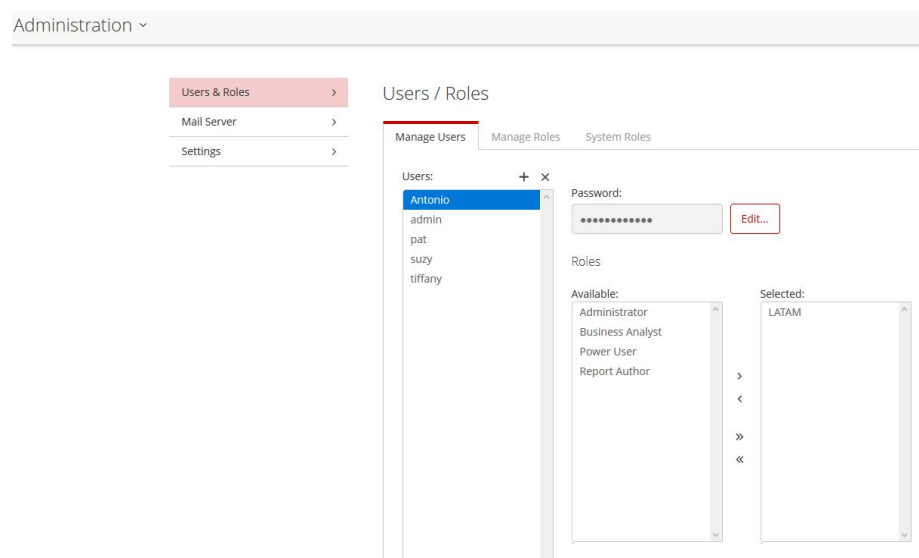


Figura 52-Configuração de utilizadores e grupos no BI Server

ANEXO A - Dados do Ficheiro LEI 2013

NFreq.	Exame	Nota	Disciplina	Época	Regime	Horário	Entrada	AEnt	AnoD	Semestre	AnoAluno		
9,5	7,7	9	ALGAV	NM	Parcial	N	Desc.	0	3	5	3		
		SMS	ALGAV	RE	Parcial	N	Desc.	0	3	5	3		
		11	10	ALGAV	C2	Parcial	N	Desc.	0	3	5	3	
10	FT	10	COMPA	NM	Parcial	N	Desc.	0	3	6	3		
		NC	FSIAP	NM	Parcial	N	Desc.	0	2	3	3		
		4,6	9	FSIAP	RE	Parcial	N	Desc.	0	2	3	3	
	1,8	7	FSIAP	C2	Parcial	N	Desc.	0	2	3	3		
		9,7	10	IARTI	NM	Parcial	N	Desc.	0	3	6	3	
		FT	NC	PESTI	NM	Parcial	N	Desc.	0	3	6	3	
15	11,6	13	SGRAI	NM	Parcial	N	Desc.	0	3	5	3		
16,1	FT	NC	ALGAV	NM	Parcial	D	Desc.	0	3	5	3		
	15,5	16	ALGAV	RE	Parcial	D	Desc.	0	3	5	3		
		14,6	12,8	14	CORGA	NM	Parcial	D	Desc.	0	3	6	3
		NF	LAPR5	NM	Parcial	D	Desc.	0	3	5	3		
13,9	FT	SMR	RCOMP	NM	Parcial	D	Desc.	0	2	4	3		
0	12,9	14	RCOMP	RE	Parcial	D	Desc.	0	2	4	3		
		SMNF	SGRAI	NM	Parcial	D	Desc.	0	3	5	3		
		14,8	15	ALGAN	NM	Parcial	N	Desc.	0	1	1	1	
0		0	MATCP	NM	Parcial	N	Desc.	0	1	2	1		
NC	FT	NC	MDISC	NM	Parcial	N	Desc.	0	1	2	1		
17,8	15,8	17	ALGAV	NM	Integral	D	Desc.	1	3	5	3		
17	12	15	ARQSI	NM	Integral	D	Desc.	1	3	5	3		
13,7	17,1	15	ASIST	NM	Integral	D	Desc.	1	3	5	3		
11		11	COMPA	NM	Integral	D	Desc.	1	3	6	3		
15,5	14,5	15	CORGA	NM	Integral	D	Desc.	1	3	6	3		
14,5	18,3	16	GESTA	NM	Integral	D	Desc.	1	3	5	3		

15	16	17	18	19	20	21	22	23	24	25	26	27
D AtrasadaAux	C/D Atrasada	AuxAno	DAnosAvançados	Res.	NDAux	NDisc	NDApAux	NDAp	PDAp	AprovDisc	NFreqArredondada	NFREQPos
0	0	3 NM	Reprov.	1	1		0				10	1
0	0	3 NM	SMS	1	1		0				0	0
0	0	3 NM	Aprov.	1	1		1			Aprovado	0	0
0	0	3 AT	Aprov.	2	2		2			Aprovado	10	1
1	1	3 AT	NC	3	3		2				0	0
1	0	3 AT	Reprov.	3	3		2			Reprovado	0	0
1	0	3 AT	Reprov.	3	3		2				0	0
0	0	3 AT	Aprov.	4	4		3			Aprovado	0	0
0	0	3 AT	NC	5	5		3			NC	0	0
0	0	3 NM	Aprov.	6	6	6	4	4	0,7	Aprovado	15	1
0	0	3 NM	NC	1	1		0				16	1
0	0	3 NM	Aprov.	1	1		1			Aprovado	0	0
0	0	3 AT	Aprov.	2	2		2			Aprovado	15	1
0	0	3 NM	NF	3	3		2			NF		0
1	1	3 AT	SMR	4	4		2				14	1
1	0	3 AT	Aprov.	4	4		3			Aprovado	0	0
0	0	3 NM	SMNF	5	5	5	3	3	0,6	SMNF	0	0
0	0	1 NM	Aprov.	1	1		1			Aprovado	15	1
0	0	1 AT	Reprov.	2	2		1			Reprovado	0	0
0	0	1 AT	NC	3	3	3	1	1	0,3	NC		0
0	0	3 NM	Aprov.	1	1		1			Aprovado	18	1
0	0	3 NM	Aprov.	2	2		2			Aprovado	17	1
0	0	3 NM	Aprov.	3	3		3			Aprovado	14	1
0	0	3 AT	Aprov.	4	4		4			Aprovado	11	1
0	0	3 AT	Aprov.	5	5		5			Aprovado	16	1
0	0	3 NM	Aprov.	6	6		6			Aprovado	15	1
0	0	3 AT	NC	7	7		6				FT	0

28	29											
SomaNDAttrasadas	TotDATrasadas	Turma	FiltroRep	Dif_FREQ_FINAL								
0	0	3NC	0	-0,5								
0	0	3NC	0									
0	0	3NC	1									
0	0	3NC	1	0								
1	1	-	0									
1	1	-	0									
1	1	-	1									
1	1	3NC	1									
1	1	*	1									
1	1	1 3NC	1	-2								
0	0	3DJ	0									
0	0	3DJ	1									
0	0	3DE	1	-0,6								
0	0	3DB	1									
1	1	2NB	0									
1	1	2NB	1									
1	1	1 3DJ	1									
0	0	1NB	1	0,2								
0	0	1NB	1	0								
0	0	0 1NB	1									
0	0	3DA	1	-0,8								
0	0	3DA	1	-2								
0	0	3DA	1	1,3								
0	0	3NB	1	0								
0	0	3NB	1	-0,5								
0	0	3DA	1	1,5								
0	0	3NB	0									

Figura 1-Dados do ficheiro Lei2013.xls