



UNIVERSIDADE DE ÉVORA
ESCOLA DE CIÊNCIA TECNOLOGIA
DEPARTAMENTO DE INFORMATICA

**Das Bases de Dados aos Sistemas de Apoio à Decisão:
Ensaio de Aplicação a uma Instituição do Ensino
Superior**

Aristidis de Jesus Ornai

Orientação:

Professora Doutora Irene Pimenta Rodrigues

Professora Doutora Lígia Maria Rodrigues da Silva Ferreira

Mestrado em Engenharia Informática

Dissertação

Évora, 2014



UNIVERSIDADE DE ÉVORA
ESCOLA DE CIÊNCIA TECNOLOGIA
DEPARTAMENTO DE INFORMATICA

UNIVERSIDADE DE EVORA
FACULDADE/ESCOLA DE CIÊNCIA TECNOLOGIA

**Das Bases de Dados aos Sistemas de Apoio
à Decisão: Ensaio de Aplicação a uma
Instituição do Ensino Superior**

Aristidis de Jesus Ornai

Orientação:

Professora Doutora Irene Pimenta Rodrigues

Professora Doutora Lígia Maria Rodrigues da Silva
Ferreira

Mestrado em Engenharia Informática

Dissertação

Évora, 2014

“Para vencer nós temos que lutar, e essa luta muitas vezes significa dispor de certa forma com os nossos estudos, para enriquecer aquilo em que acreditamos. O nosso ponto de vista, a nossa confiança e a nossa personalidade acima de tudo. E se nós não lutamos para valer, acabamos por perder o nosso próprio rumo. No entanto, para conseguirmos manter essa linha de conduta, temos que lutar muito”

»Aristidis Jesus«

“A morte do homem começa no instante em que ele desiste de aprender”

»Albino Texeira«

Agradecimentos

Gostaria de agradecer a todos aqueles que me apoiaram e contribuíram de forma direta e indireta para a conclusão deste Mestrado, e em particular:

- A **Deus** por me apoiar nos momentos difíceis, dar-me força interior para superar as dificuldades, mostrar o caminho em tempos de incerteza e encontrar-me em todas as minhas necessidades.
- À **Professora Doutora Irene Rodrigues** e **Professora Doutora Lígia Ferreira** pelas orientações, acompanhamento, incentivo e permanente disponibilidade demonstrada durante este projecto até a conclusão.
- Ao responsável da Universidade Nacional Timor Lorosae (UNTL), **Professor Doutor Aurélio Guterres** e toda a sua estrutura deram-me esta grande oportunidade e todos os suportes necessários até ao fim do curso.
- Ao decano da Faculdade de Engenharia Ciências e Tecnologia da UNTL, **Professor Gabriel de Sá** e toda a sua estrutura assim como os chefes de cada Departamento que me autorizou a licença de recolher os dados da Faculdade.
- Ao responsável da Fundação das Universidades Portuguesas (FUP) **Doutora Mónica Pimentel**, pela sua grande responsabilidade e coragem que me deu durante a minha estadia em Portugal.
- À minha família, e em particular aos meus pais, pelo permanente estímulo e natural apoio demonstrado ao longo da realização desta dissertação e trabalhos relacionados.
- Aos meus amigos que me ajudaram, apoiaram e incentivaram tornando mais fácil a conclusão do ano curricular.
- Aos meus companheiros estudantes Timorenses, à **Associação dos Estudantes Timorenses em Évora (ASET-Évora)** e todos os outros que de forma directa e indirecta contribuíram durante minha estadia em Évora.

RESUMO

Actualmente no mundo das Tecnologias da Informação tem-se vindo a ganhar mais conhecimento da Tecnologia de Data Warehouse (DW) e cada vez mais aprofundadamente do que era nas gerações passadas. DW é um sistema de coleção de uma variedade de dados necessários, seja na forma de dados históricos ao longo de cinco anos ou seja, dados actuais. DW é geralmente apoiado por um modelo conceptual de dados chamado modelo multidimensional que pode ser usado para exibir dados de diferentes dimensões de análise e relatórios. Os dados são uma mais valia importante nas organizações e instituições em que são usados para fazer a tomada de decisão. O processo da recolha de dados pode ser realizado em vários lugares, por exemplo, bases de dados, aplicações e tecnologia de DW.

O sistema de DW foi desenvolvido através da plataforma de *Business Intelligence* (BI) da Microsoft. Aplicando as tecnologias de modelação multidimensional é possível, quando após a selecção, preparação e limpeza dos dados extraídos dos sistemas transaccionais da UNTL. Portanto os dados são carregados para DW e procederam às respectivas análises através de tecnologias *Online Analytical Processing* (OLAP).

Este estudo tem como objetivo sensibilizar no interesse e vantagem do sistema da construção, implementação, gestão para actividades académicas que podem ser usados para apoiar a tomada de decisão a nível universitário. Nesta dissertação, proposto o padrão de desenvolvimento baseada ao estudo de caso na UNTL. A arquitetura que propoe vai ajudar a desenvolver o sistema de tomada de decisões de qualquer instituição académica.

Palavra-chave: Base de Dados, ETL, Sistemas de Apoio à Decisão, Sistemas de Data Warehouse e Business Intelligence.

From Databases to Decision Support Systems: An Application to the University of Timor Leste

ABSTRACT

Current scenario of Information Technology in the world has steadily gaining more knowledge of Data Warehouse (DW) Technology day by day. DW is usually supported by a conceptual data model called multidimensional model that can be used to display data from different dimensions of analysis and reporting. DW is a system of collection of a variety of data required, whether in the form of historical data over 5 years or current data. The data are an important input in organizations and institutions that are used to make the decision. The process of data collection can be accomplished in several places, for example, databases, applications, and DW technology.

A system of DW was developed by the Business Intelligence (BI) platform of Microsoft. Implementing technologies of multidimensional modeling is possible after the selection, preparation and cleaning of extracted data from transactional systems UNTL only. Therefore, data is loaded to DW and carried out their analysis technologies through Online Analytical Processing (OLAP).

This thesis aims to raise awareness, interest and advantage of building system along with its implementation and management in academic activities that can be helpful in decision making at the university level. In this thesis, proposed framework is developed on the basis of case study of University of East Timor. The proposed architecture will be help to develop the decision making system of any academic institution.

Keywords: Database, ETL, Decision Support Systems, Data Warehouse Systems and Business Intelligence.

Índice

Agradecimentos	i
RESUMO	ii
ABSTRACT	iii
LISTA DE FIGURAS	vi
LISTA DE TABELAS	vii
LISTA DE ACRÓNIMOS E ABREVIATURAS	viii
Capítulo 1 Introdução e Objetivos	1
1.1 Introdução e Motivação	1
1.2 Objetivos do Trabalho	3
1.3 Metodologia	3
1.4 Organização do Documento	4
Capítulo 1 – Introdução e Relevância do Tema	4
Capítulo 2 – Fundamentos Teóricos	4
Capítulo 3 – Ferramentas Utilizadas Para O Caso Prático	5
Capítulo 4 – Estudo de Caso Prático	5
Capítulo 5 – Conclusão e Trabalhos Futuros	5
1.5 Cronograma do Trabalho	5
Capítulo 2 Fundamentos Teóricos	7
2.1 Conceitos Básicos	7
Online Transactional Processing (OLTP) vs Online Analytical Processing (OLAP)	8
2.2 As Principais Características de Uma DW	10
Orientado por assunto	10
Integrado	10
Não Volátil	11
Variáveis no tempo / histórico	12
2.3 Tecnologia de DW	13
2.3.1 Modelo de dados	13
Tabelas de Dimensão	13
Tabelas de Factos	14
O Modelo Estrela (star schema)	15
O Modelo Floco de Neve (Snow Flake)	16
Modelo Normalizado	18
2.3.2 Arquitetura de DW	18
Sistemas Operacionais	19
Aquisição de dados (Data acquisition)	19
Metadados	20
Armazenamento de dados (Data storage)	20

Acesso aos Dados.....	21
Data Marts	22
Business Intelligence(BI)	22
Concepção do Data Warehouse.....	23
2.3.3 Extração Transformação e Carregamento	23
Extração	24
Transformação	25
Carregamento	26
Capítulo 3 Ferramentas Utilizadas Para O Caso Prático	27
3.1 Microsoft SQL Server	27
3.2 SQL Server Analysis Services (SSAS).....	27
3.3 Microsoft Office Excel	28
3.4 Ferramentas Front End	28
3.5 Microsoft SQL Server Report Builder	28
Capítulo 4 Estudo de Caso Prático	29
4.1 Problema.....	29
4.2 Base de Dados.....	29
4.3 Análise do Processo de Negócio	32
Processo de Negócio	32
Matriz em Bus	33
4.4 Camada de Visualização de Dados.....	34
4.5 Desenvolvimento do Modelo Dimensional.....	34
4.6 Análise Multidimensional e Relatórios	41
Capítulo 5 Conclusões e Trabalhos Futuros.....	46
Conclusões	46
Bibliografia	47
Anexos.....	49

LISTA DE FIGURAS

Figura 1 Plano de Trabalho.....	5
Figura 2 Integração de Dados (Inmon, 2002)	11
Figura 3 Não Volátil (Inmon, 2002)	12
Figura 4 Variáveis no Tempo/Histórico (Inmon, 2002)	13
Figura 5 Exemplo Modelo de Estrela (<i>Star Schema</i>). Fonte: Elaboração própria.....	16
Figura 6 Exemplo Modelo Floco de Neve (<i>Snow Flake</i>). Fonte: Elaboração própria.....	17
Figura 7 Arquitetura de DW. Fonte: Elaboração própria	18
Figura 8 Extração, Transformação e Carregamento de Dados (ETC). Fonte: Elaboração própria.....	24
Figura 9 Dados dos Alunos de Informática em Excel	30
Figura 10 Conversão de Dados com Linguagem SQL	31
Figura 11 Exemplo Dados dos Alunos no SQL Server 2012	31
Figura 12 Diagrama de Base de Dados Relacional.....	32
Figura 13 Modelo Dimensional para o Processo de Negócio Inscrição na Disciplina	39
Figura 14 Modelo Dimensional para o Processo de Negócio Avaliação_Pauta do Aluno.....	40
Figura 15 Cubo de Dados Gerado no BIDS.....	41
Figura 16 Browse do Cubo de Dados para relatório.....	42
Figura 17 Inscrição do Aluno na Disciplina por Ano Letivo.....	43
Figura 18 Gráfico de Inscrição do Aluno na Disciplina por Ano Letivo.....	44
Figura 19 Pautas de Avaliação do Aluno.....	45
Figura 20 Gráfica de Avaliação do Aluno por Pauta.....	45

LISTA DE TABELAS

Tabela 1. Comparação entre o sistema de OLTP e OLAP (Jiawei Han, 2001).....	9
Tabela 2 Matriz em BUS	33
Tabela 3 Descrição de Dimensão Aluno.....	35
Tabela 4 Descrição de Dimensão Curso	35
Tabela 5 Descrição de Dimensão Disciplina	36
Tabela 6 Descrição de Dimensão Professor.....	36
Tabela 7 Descrição da Dimensão Pauta.....	37
Tabela 8 Descrição da Dimensão AnoLectivo.....	37
Tabela 9 Descrição da Dimensão Reingresso	38
Tabela 10 Descrição de Facto Inscrição na Disciplina	38
Tabela 11 Descrição de Facto Pauta	39

LISTA DE ACRÓNIMOS E ABREVIATURAS

Ao longo deste documento utilizaram-se os seguintes acrónimos e abreviaturas:

ASCII = American Standard Code for Information Interchange

ASET-Évora = Associação dos Estudantes Timorenses em Évora

BD = Base de Dados

BI = Business Intelligence

BIDS = Business Intelligence Development Studio

DW = Data Warehouse

DBA = Data Base Administration

DMX = Data Mining Extension

DTS = Data Transformation Service

ETL = Extraction Transformation and Loading

ETC = Extração Transformação e Carregamento

FUP = Fundação das Universidades Portuguesas

GB = Giga Byte

H/M = Homem/Mulher

MB = Mega Byte

M/F = Masculino/Feminino

MDX = MultiDimensional eXpressions

OLTP = Online Transactional Processing

OLAP = Online Analytical Processing

ODS = Operational Data Source

SQL = Structure Query Language

SGBD = Sistema de Gestão de Base de Dados

SGBDR = Sistema de Gestão de Base de Dados Relacional

SSAS = SQL Server Analysis Services

SP1 = Service Pack 1

TB = Terra Byte

UNTL = Universidade Nacional Timor Lorosae

XML/A = eXtensible Markup Language for Analysis

Capítulo 1 Introdução e Objetivos

1.1 Introdução e Motivação

Actualmente no mundo das Tecnologias da Informação tem-se vindo a ganhar mais conhecimento da Tecnologia de DW e cada vez mais aprofundamente do que era nas gerações passadas. Com a evolução e a globalização tecnológica em avanço a um crescimento exponencial, a informação é cada vez mais importante para pessoas e instituições. Nos últimos anos, o sector comercial tornou-se mais competitivo, de tal modo que as empresas e os clientes são extremamente ágeis e exigentes. O ciclo de vida de desenvolvimento dos produtos torna-se cada vez menor, obrigando a que as empresas e os gestores tenham necessidade de identificar problemas, riscos, oportunidades ao tomarem decisões operacionais e estratégicas em intervalos decrescentes de tempo. Dada esta situação, a competitividade de uma organização passa a ser determinada, em grande parte, pela sua estratégia de negócio. Essas decisões devem ser baseadas em informações atualizadas, completas e de elevada qualidade em diversas aplicações.

Nas décadas recentes, os sistemas DW surgiram como uma solução para os problemas de apoio à decisão. Os gestores e executivos necessitam de recursos computacionais que fornecem suporte para apoio ao processo decisório, sobretudo nos níveis tático e estratégico das empresas. A informação tornou-se o bem mais valioso e o mais interessante é que esta informação está disponível nas próprias organizações e instituições. Os dados para a obtenção de informação estão armazenados em bases de dados operacionais, arquivos pessoais e documentos. O grande desafio é saber como tornar estes dados úteis para o sistema de apoio a decisão, transformando dados em informação e esta em ações que melhorem o processo de tomada de decisão.

Para aceder a uma base de dados gigantesca e dispersa e conseguir concretizá-la numa fonte de informação viável, é necessário transformar e organizar os dados para a execução de informações estratégicas úteis. As decisões fazem parte da informação que está disponível e são tanto mais eficazes quando se torna necessário reestruturar e repensar os sistemas de instituições com o objetivo de na melhor forma disponibilizar a informação relevante para cada tipo de decisão. Esse tipo de informação que circula numa instituição é específico para cada nível da estrutura institucional. Desse modo, um gestor operacional necessita de informação para tratar e armazenar todo processo de transações diárias e necessários aos sistemas transacionais, pois a nível estratégico, a informação

deverá ser proveniente de diversas fontes sob um ponto de vista analítico e sistemas de suporte à decisão.

A crescente evolução tecnológica e crescente utilizador de computadores ligados em rede, permite que médias e grandes empresas, instituições e organizações utilizem sistemas informáticos para realizar os seus processos úteis, o que com o passar do tempo gerou uma quantidade enorme dos dados relacionados. Esses dados armazenados num ou mais sistemas operacionais de uma empresa, instituição e organização identificam-se como um recurso, mas geralmente é muito raro servirem como recurso estratégico.

As decisões são normalmente tomadas com base na experiência dos administradores da instituição mas poderiam também ser baseadas em factos históricos armazenados pelos diversos sistemas de informação, utilizados pelas instituições. Sob esta condição surgiu o *Data Warehouse*, como um desafio para integrar os dados, eliminadas as redundâncias e identificadas informações iguais que pudessem estar representadas sob formatos diferentes em sistemas distintos. Geralmente é utilizado em grandes corporações e coleções aos dados de sistemas transacionais diversos.

No conceito principal de DW a importância mais relevante são os dados como fontes do controlo operacional e da estratégia da tomada a decisão. Este sistema tornou-se numa peça muito útil, dado que o processo de armazenamento de dados tem cada vez mais vindo a ganhar força devido à evolução do hardware e software.

Hoje em dia as instituições de ensino superior são obrigadas a ser mais competitivas, utilizando todos os recursos disponíveis. Além de recursos de infraestrutura, instalações e pessoas, os sistemas de informação são um dos recursos que podem ser usados para melhorar a competitividade. Os sistemas de informação podem ser utilizados para adquirir, processar e disseminar informações para apoiar as operações diárias, bem como apoiar a tomada de decisões estratégicas. O desafio frequentemente encontrado é o facto de o sistema ser perfeito e as informações resultantes apresentado incompletas ou incorretos.

Numa instituição do ensino superior é necessária servir todas as partes interessadas, tais como o estudante. Os exemplos de serviços prestados aos estudantes incluem pagamentos relativos à administração, a administração das inscrições nas disciplinas, administração da faculdade, entre outros (Caldeira, 2008). Um sistema integrado de informação é certamente necessário para acelerar as atividades de serviços. Dados das atividades académicas de ano para ano vão continuar a crescer e gerir bancos de dados cada vez maiores. A fim de não sobrecarregar os sistemas de processamento de transações existentes, os dados mais antigos precisam de ser transferidos para um DW. Um DW irá fornecer um relatório que é dinâmico e pode ser visto a partir das várias dimensões. O relatório que foi gerado também terá a capacidade para mais detalhes e informação mais resumida. Sem o

DW, os relatórios gerados serão estáticos de acordo com os dados das aplicações que estão no nível dos sistemas de informação de gestão.

1.2 Objetivos do Trabalho

O estudo proposto por este trabalho pretende-se impressionar no interesse e vantagem de construção, implementação, gestão e utilização do sistema de DW no ensino superior da UNTL. De uma forma geral, tem alguns objetivos a considerar neste trabalho, dos quais foram:

- Desenvolver um sistema transaccional;
- Projetar e desenvolver um sistema de DW;
- Aplicar as tecnologias de modelação multidimensional;
- Processar informação dos sistemas transaccionais para sistema de análise DW através de tecnologias OLAP;
- Produzir e disponibilizar relatórios que possibilitem extrair informação coerente.

Esses objetivos são mesmo importantes para analisar as atividades académicas para apoiar a tomada de decisão a nível universitário. Podemos dizer que a informação que se passa na instituição será mais útil e bem decidido de maneira como foi utilizada, criando assim, a possibilidade de melhorar as condições da instituição tornando a mais competitiva. Um dos problemas que existe nas instituições em termos da informação é saber como organizar os dados a fim de transformá-los em informações úteis, relevantes e estratégicas para a tomada a decisão. Dentre as possíveis opções de ferramentas de gerenciamento e organização de informações, o DW aparecem como uma das aplicações mais importantes. Realmente, o DW possibilita explorar os dados transformando-os em informações necessárias à tomada de decisão, podendo assim gerar melhores resultados para uma instituição.

Durante a elaboração deste trabalho verificou-se a falta de informações em condições, devido aos dados recolhidos serem da UNTL. Por fim, considera-se que esta informação seria importante para implementar um sistema de transaccção para avaliar as informações mais necessárias e assim, prover informações de maneira organizada, selecionada e suficiente o que é certamente uma contribuição essencial à instituição. No entanto, o objetivo de DW é fornecer a informação certa, para a pessoa certa, no momento exato.

1.3 Metodologia

Devido à singularidade do complexo sistema de informação da UNTL, foi necessário definir uma abordagem metodológica, que embora baseada nas tradicionais, possuísse características

próprias para se adequar neste contexto. A estrutura básica desta metodologia corresponde ao fluxo de atividades que são estudo dos dados, elaboração do modelo relacional, elaboração do modelo dimensional, avaliação dos modelos, processo de carga, desenvolvimento de relatórios e finalmente divulgados dos relatórios aos utilizadores.

Um ponto incluído na metodologia foi a transformação dos dados, já que os dados originais não estão no modelo relacional. Inicialmente foi criado um modelo relacional no SQL Server e foi feito inserido dos dados de originais obtidos através de uma folha de cálculo para o modelo em questão. Os dados recolhidos foram importantes, dado que são fundamentais para este tipo do projeto de tese correspondendo aos dados de professores, alunos, disciplinas e cursos. Esses dados foram obtidos no serviço académico e nas direções da faculdade de Engenharia do ensino superior da UNTL. Contudo os dados são suficientes para a realização da parte prática do trabalho.

Na etapa final, que o desenvolvimento dos relatórios geridos por objetivo de garantir a qualidade de informação produzida. Inicialmente desenvolverá alguns modelos de relatórios importantes baseado ao processo de negócio e ao fim publicados para o utilizador final.

Uma etapa persistente em qualquer projeto é a definição das ferramentas para o desenvolvimento. O processo de escolha abrange vários aspetos, dentre eles, robustez para suportar o volume de dados, existência de casos de sucesso implementados na ferramenta e um custo acessível. Considerando estes aspetos, foi adotado um conjunto de ferramentas de apoio do Microsoft Visual Studio 2012. Na execução de relatórios foi utilizada a ferramenta do Report Builder 3.0 que possui algumas funcionalidades que atendem características específicas do projeto, tais como, a criação dos ditos relatórios.

1.4 Organização do Documento

Este documento de dissertação encontra-se dividido em seis capítulos:

Capítulo 1 – Introdução e Relevância do Tema

Este capítulo descreve uma apresentação do histórico do DW desde o início até à atualidade. Pretende também descrever as motivações e objetivos envolvidos ao DW em como será útil à tomada de uma decisão.

Capítulo 2 – Fundamentos Teóricos

Pretende-se enquadrar historicamente os sistemas de DW e fornecer definições para estes sistemas de acordo com alguns autores. Por outro lado, também descreverem-se as características de um DW e o modelo de arquitetura de dados, bem como o seu processo de desenvolvimento.

Capítulo 3 – Ferramentas Utilizadas Para O Caso Prático

Descreve as ferramentas utilizadas para o desenvolvimento da aplicação de DW.

Capítulo 4 – Estudo de Caso Prático

Descreve uma parte prática desta dissertação após ter explicação da parte teórica.

Capítulo 5 – Conclusão e Trabalhos Futuros

Pretende-se efetuar uma conclusão e o trabalho futuro que será um alvo de criticar e sugerir coisas úteis nesta dissertação.

1.5 Cronograma do Trabalho

O plano proposto para a implementação e desenvolvimento ao longo deste projeto, pode ver-se na Figura 1.

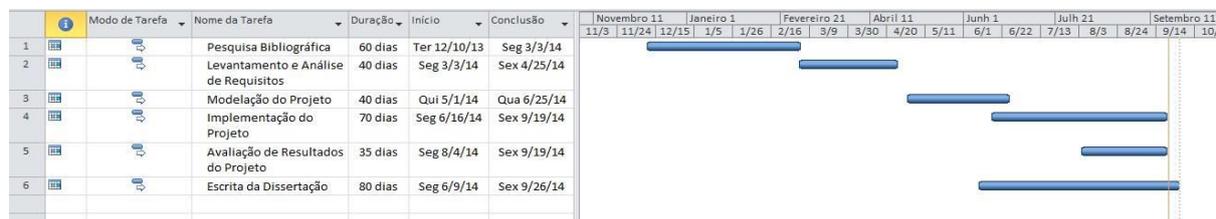


Figura 1 Plano de Trabalho

As descrições dos planos de trabalho ao longo deste projeto:

Fase 1: A pesquisa de bibliografia sobre o tema, bem como análise pormenorizada do estado da arte no que diz respeito à implementação e desenvolvimento do sistema de DW.

Fase 2: Esta segunda fase tem por objetivo a análise e especificação dos requisitos funcionais do DW. Inicialmente, efetua uma análise mais profunda dos sistemas operacionais existentes na instituição da UNTL, que deverão ser então, projetados o modelo conceptual e lógico do DW.

Fase 3: É útil perceber as fontes e bases de dados operacionais, a escolha da arquitetura e infraestrutura da tecnologia a implementar. Nesta terceira fase pretende-se efetuar a modelação multidimensional que constituirá o DW. Esta modelação do projeto, conjuntamente com uma análise profunda das fontes de dados, tanto ao nível de sistemas que os originam como dos próprios dados em si, devera ser feita de modo a poder elaborar uma estratégia para o processo de ETL dos dados.

Fase 4: Considera-se como a continuação dos planos anteriores na estrutura do projeto. Obviamente é feito por desenvolvimento dos componentes de suporte ao processo de DW destacando-se os processos de ETL dos dados. Evidentemente, será realizar à análise multidimensional, as interrogações em SQL, multidimensional aos cubos de dados e a transformação dos resultados para os relatórios.

Fase 5: Pretende-se avaliar os resultados obtidos através da ferramenta utilizada nos relatórios.

Fase 6: Estabelecimento da escrita da dissertação.

Capítulo 2 Fundamentos Teóricos

2.1 Conceitos Básicos

Atualmente no mundo da tecnologia, há um grande volume de dados nas empresas mas tenham falta de informações significativas. Este é o caso devido à extensa utilização de sistemas de informação específicos, contribuindo para a pouca integração dos dados e a existência de ambientes não é muito adequados para o tratamento analíticos dos dados. O ambiente de DW atuar a necessidade em questão, promovendo a integração dos dados dos sistemas de informação espalhados pela organização e gerando um ambiente de consultas com informações significativas, neste sentido análises que ajudam no processo de tomada de decisão.

Existem vários conceitos de DW propostos por diferentes perspectivas de autores. Assim apresentam-se dois dos principais conceitos que são considerados mais importantes.

Primeiro, o conceito defendido pelo professor Ralph Kimball que é um dos mais conceituados precursores de conceitos de DW. Este autor refere que um DW não é mais do que a união de um conjunto de *Data Marts*. Um *Data Mart* pode ser definido como um DW de pequena capacidade que abrange uma determinada área ou departamento, oferecendo informações mais detalhadas sobre um determinado assunto em questão. O paradigma que Ralph Kimball defende baseia-se no facto da informação ser guardada utilizando o modelo dimensional (Raph Kimball, 1998) .

Segunda, outra das definições mais difundidas de DW foi dada por Bill Inmon que descreve “DW como uma coleção de dados integrados, orientados, por assunto, não voláteis, variáveis no tempo, e que fornecem suporte ao processo de apoio à decisão ” (Inmon, 2002). O paradigma deste autor difere um pouco do paradigma de Kimball e defende que um DW é apenas uma parte de todo o processo de *BI*. Uma organização possui por um DW de onde os *Data Marts* extraem a informação, sendo a informação guardada na terceira forma normal. Ambos os paradigmas são válidos na medida em que representam filosofias diferentes de DW.

Uma definição mais expansiva para DW também inclui o *BI* que contém ferramentas para extração, transformação e carregamento de dados no repositório por motivo de obter e gerir os seus metadados. Desde já, a existência de *BI* no mercado de ferramentas de DW é considerado como as melhores ferramentas mais sofisticadas de que foram sendo desenvolvidas para apoiar a estrutura do DW e da sua utilização. Com base nestas definições, podem-se dizer de forma resumida, que o DW é uma base de dados de integração especial, que serve para dar suporte às altas gestores das empresas para análise de tendências históricas dos seus produtos, funcionários e clientes. Sabia que no DW os

dados são extraídos de múltiplas fontes de dados internas ou externas da empresa e organizações que no fim integrados num conjunto de ficheiro. O DW é construído para que tais dados possam ser armazenadas e acedidos de forma que não sejam limitados por tabelas e linhas estritamente relacionais.

Assim elabora-se uma solução de DW significativo que descreve como um repositório unificado e consolidado da informação da organização, com um determinado histórico; os processos de extração, transformação e carregamento de informação, proveniente de diferentes sistemas operacionais; os processos de limpeza de dados; os repositórios departamentais (*Data Marts*) e as aplicações no cliente para que podem estabelecer a interação de consultas a informação do repositório (Daniele Del Bianco Hokama, 2004).

Online Transactional Processing (OLTP) vs Online Analytical Processing (OLAP)

É de fundamental importância para este trabalho explicitar as diferenças existentes entre essas duas metodologias de armazenamento de dados, para permitir melhor compreensão de seus objetivos. A principal diferença se caracteriza pelo DW ser um ambiente informacional, enquanto a base de dados convencionais das organizações, como se localizados em suas filiais, serem ambientes operacionais. No DW existem dois processos conhecidos e principais aos desenvolvedores dessa aplicação.

Conhecemos primeiro (Inmon, 2005), os sistemas de processamento de Transação On-line ou OLTP que são sistemas operacionais de registro que capturam as transações diárias de uma empresa ou organização. Estes são dados responsáveis pela persistência dos dados que os aplicativos da empresa ou organização manipulam. Esse tipo de base é utilizado em sistemas que registam pequenas transações como *Insert*, *Update* e *Delete* realizadas em tempo real e que ocorrem constantemente e de forma rápida. Enquanto neste sistema operacional os dados podem ser alterados. Por não salvar histórico dos dados, isso não o qualifica como base de dados ideal para ajudar na tomada de decisões. Também é uma base que necessita de *backup* regularmente, pois caso o banco de dados seja perdido os dados não podem ser reaproveitados.

Por fim a estes sistemas que os dados serão retirados e serem transformados em informação para a empresa ou organização. Enquanto por exemplo na empresa, se tivéssemos um sistema *on-line* de vendas e a empresa resolvesse começar a executar consultas imensas e demoradas, o sistema ficaria muito lento e provavelmente, não conseguiria registar a venda por estar ocupado processando a consulta.

O segundo processo conhecido é o OLAP disponibilize uma alternativa diferente. Este sistema analítico é constituído por um conjunto de tecnologias especialmente focados em dar suporte ao processo de decisão para as analistas, os gestores e os executivos. Também proporciona uma visão dos dados orientados à análise, cálculos, consultas e além de uma navegação rápida, consistente e flexível (Inmon, 2005). Este sistema de OLAP tem uma forma de receber os dados do OLTP para que possa realizar as análises. Essa carga de dados acontece conforme a necessidade da empresa ou organização. Sendo assim, um sistema para tomada de decisões, não se realizam transações como aquilo que o sistema OLTP faz *Insert*, *Update* e *Delete*, pois a sua finalidade são as consultas. Este sistema possui dados atuais e históricos e não há necessidade de *backups* regularmente, sendo que ele possui informações do OLTP. Por fim esses dados são armazenados em DW e necessita das partes de analistas e gestores a tomar decisões.

Alguns autores (Jiawei Han, 2001) descrevem as diferenças entre os sistemas de OLTP e o ambiente de DW (OLAP), como na Tabela 1.

Tabela 1. Comparação entre o sistema de OLTP e OLAP (Jiawei Han, 2001)

Característica	OLTP	OLAP
Propósito	Processamento operacional	Processamento informacional
Orientado a	Transações	Análises
Funções	Operações dia a dia	Suporte à decisão
Base de dados (BD)	Orientado a aplicação	Orientado a assunto
Dados	Correntes, detalhados	Consolidados, históricos
Unidade de trabalho	Transação	Consultas complexas
Acesso	Leitura, incluir e excluir	Leitura
Foco	Entrada de dados	Saída de informações
Número de registros	Dezenas	Milhares
Medida de performance	Tempo processamento de transação	Tempo de processamento de consulta
Utilizador	Administrador de base de dados (DBA), base de dados profissional	Gestores, analistas e executivo

Tamanho de dados	100MB até GB	100GB até TB
Atualização	Contínuo (tempo real)	Periódica (em batch)
Redundância	Não ocorre	Ocorre
Granularidade	Detalhados	Detalhados e resumidos
Histórico	60 a 90 dias	5 a 10 anos

Como referimos na diferença entre os dois na Tabela 1, é difícil de combinar as capacidades do sistema OLAP e OLTP num só sistema. Em princípio o modelo dimensional que utilizada no DW é muito mais efetivo para a consultas comparado com o sistema OLTP no modelo relacional de base de dados. O modelo dimensional de DW não é adequado para os sistemas de OLTP, essencialmente devido à redundância e da perda de integridade referencial dos dados.

2.2 As Principais Características de Uma DW

Há quatro propriedades que caracterizam os dados armazenados no DW que é definida por um dos criadores DW (Inmon, 2002).

Orientado por assunto

A orientação por assunto é uma característica mais forte que incide sobre todo o DW. É organizado em torno de assuntos importantes que devem focar ao máximo o negócio da empresa, tais como por exemplo, cliente, produto e vendas. Cada tipo de organização tem seu próprio conjunto de assuntos de informações pertinentes a área em questão. São por isso focados na modelação e análise de dados para quem toma decisões, em vez de operações diárias e processamento de transações (Inmon, 2002). Os DW são selecionados, isto é, fornecem uma visão simples sobre questões de um tema particular através da exclusão de dados que não são importantes no suporte ao processo de decisão. Pelos contrários, em ambientes operacionais as aplicações contêm dados necessários à satisfação imediata dos requisitos funcionais que podem ou não ser utilizados no processo de decisão

Integrado

O aspeto mais importante no armazenamento de dados é a integração de dados. Um DW é construído ou alimentado por integração de múltiplas fontes distintas de dados. São aplicadas

técnicas de limpeza de dados e técnicas de integração de dados. A integração de dados, provenientes de sistemas operacionais, efetua-se nos mais variados níveis, na estrutura consistente de códigos, na forma consistente das variáveis, na conversão de nomes, atributos físicos e assim por diante. Cada *designer* da aplicação teve rédea livre para fazer a suas decisões de design próprios, que com o resultado de qualquer aplicação é muito diferente de qualquer outra aplicação. Por exemplo, considere-se sexo como um elemento de dado. Uma aplicação pode codificar sexo como M/F, outra como 1/0 e uma terceira como H/M (Inmon, 2002).

Conforme Figura 2 os dados são inseridos nos DW devem estar consistentes entre si em termos de nomes, formatos e unidades de medida. Eles são convertidos para um estado uniforme, ou seja, sexo é codificado apenas de uma forma. No processo de integração dos dados também pode ser necessário corrigir dados que estejam inconsistentes na origem, devido à não integração dos sistemas transacionais que fornecem para um estado uniforme de modo a permitir a carga no DW. Na Figura 2 ilustra-se a interação que ocorre quando os dados passaram a partir do ambiente de aplicação orientada operacional para o DW.

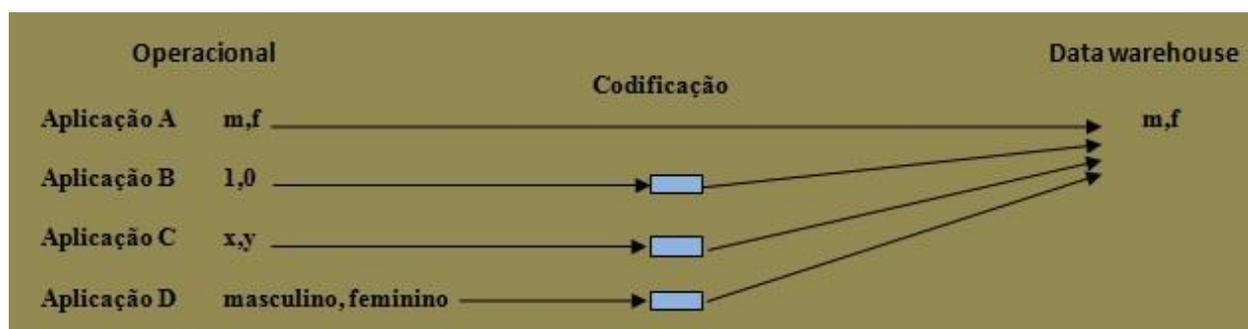


Figura 2 Integração de Dados (Inmon, 2002)

Não Volátil

No sistema operacional essa característica não volátil permite diversas operações de atualização dos dados do assunto normal como acrescentar, substituir e apagar. Pelo contrário que no DW, a carga de dados e as consultas são as duas principais operações. Alterações nos dados são pontuais e seguem regras muito rigorosas, de modo a não permitir que a credibilidade dos dados entre em cheque. Após o desenvolvimento do DW, esse se encontra vazio. A primeira carga é feita por extração dos dados completos do sistema fonte. Esses dados passam por processos de validação e integração onde sofrem alterações. Esses dados limpos e validados são carregados nas tabelas de factos e dimensões. A quando chegar ao momento de uma nova extração, o mais comum é que sejam

extraídos todos os dados novamente. Uma questão que deve ser levada em consideração é que os dados extraídos passam por um processo de Extração, Transformação e Carregamento (ETC), onde nas fases E T, os dados sofrem alterações. Após serem integrados e transformados, os dados são carregados em bloco para o DW que são estáticos, para que estejam disponíveis aos utilizadores para acesso.

Ao contrário com ambiente operacional que normalmente os dados são atualizados registro a registro, em múltiplas transações. Esta volatilidade requer um trabalho considerável para assegurar integridade e consistência através das atividades de *rollback*, recuperação de falhas, *commits* e bloqueios (Inmon, 2002). Um DW não requer este grau de controlo típico dos sistemas orientados a transações.

Na Figura 3 ilustra-se não volátil de dados e mostra que os dados operacionais são regularmente acedidos e manipulados uns registro de cada vez.

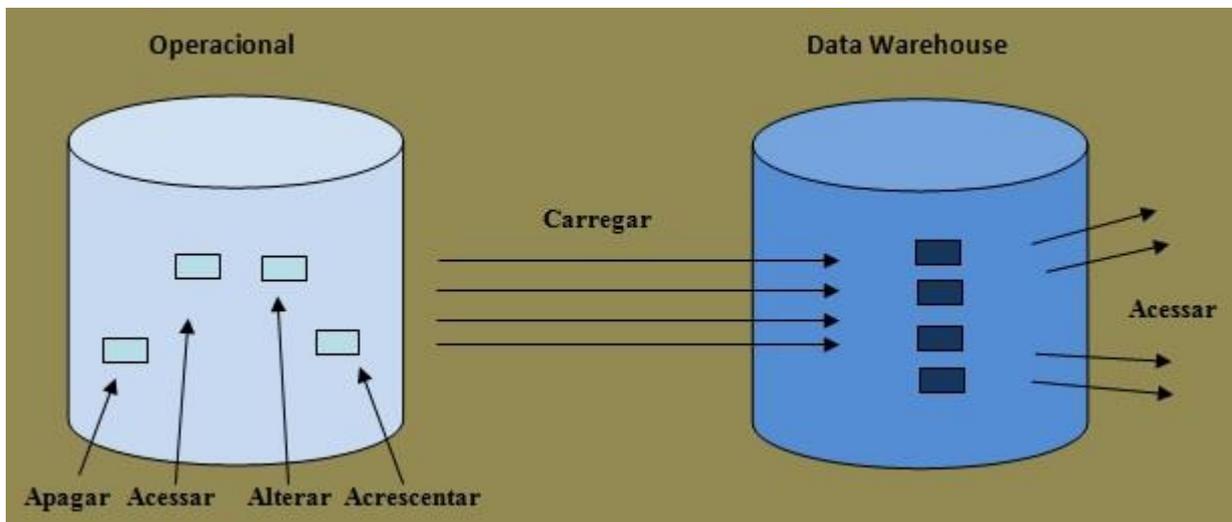


Figura 3 Não Volátil (Inmon, 2002)

Variáveis no tempo / histórico

Essa última característica salientada no armazenamento de dados implica-se a cada unidade de dados no DW é preciso como de algum momento no tempo que significa predominantemente o DW deve manter históricos dos dados. Esses históricos serão maiores que os existentes nos sistemas transacionais. Conforme na Figura 4, é comum que nos sistemas transacionais os históricos não passem de três meses, enquanto no DW pode ser de dez anos. Num ambiente operacional os dados são exatos no momento em que são acedidos, ou seja, a base de dados operacional contem a informação atual. Um DW tem geralmente diversas fontes de dados operações distintas, o que

significa aí acedido podem já não ser os mais atualizados, pois podem já ter sido modificados no sistema operacional e o refrescamento do DW ainda não ter sido efetuado. Assim, o DW representa os dados sobre um horizonte de tempo representado pelo âmbito operacional é mais curto (alguns dias).

Os dados no DW refletem situações ao longo tempo, ou seja, são dados históricos. Os dados são regularmente inseridos de modo a manter o registo histórico das operações ao longo da sua existência. Existe portanto a noção de *timestamp* e de data de validade do registo. Isto permite que as consultas possam ser executadas para recuperar o estado da informação num dado período de tempo (Inmon, 2002).

Normalmente um horizonte de tempo na Figura 4 representado num ambiente que é coletivo para os dados encontrados dentro de um DW. Todas estas características conduzem a um ambiente muito diferente do ambiente operacional. O sistema operacional contém dados atuais, enquanto o DW contém dados não só atuais, mas também dados do histórico para serem utilizados na análise e tomada de decisão. Em todos os DW suportados uma variável muito importante que se chama a dimensão do tempo. Essa dimensão existe os atributos para análise de vários processos de negócio contém vários valores de tempo, por exemplo, diários, semanais, mensais e anuais.

A Figura 4 ilustra a variação de tempo dos dados entre os sistemas operacionais e o DW.



Figura 4 Variáveis no Tempo/Histórico (Inmon, 2002)

2.3 Tecnologia de DW

2.3.1 Modelo de dados

Tabelas de Dimensão

O conceito básico da tabela de dimensão são informações descritivas relacionadas aos processos de negócio, ou seja, são as informações que vão conceituar o negócio, também definido como uma

coleção de atributos textuais ou conjuntos de elementos que são correlacionados entre si. Essas tabelas de dimensão são formadas de vários atributos pelo seu objetivo de definir regras de agrupamento e filtros para consultas em uma tabela de facto. Os atributos nas tabelas de dimensão possuem um papel importante do DW, por serem a origem de todas as restrições e rótulos de relatórios, os atributos são fundamentais para que o DW seja utilizado e compreendido (Ralph Kimball, 2002).

A cerca disto, vimos por exemplo os tipos da tabela de Aluno, tabela de Data, tabela de Docente, etc. os dados que estão armazenados nas dimensões carregadas do banco de dados de origem e são utilizados mesmo para realização de consultas facilitando a necessidade do DW. Cada tabela dimensional deve incorporar múltiplos campos de atributos contendo textos e códigos que descrevam melhor a chave. Os campos de atributos são usados para delimitar ou filtrar o conteúdo da dimensão. As tabelas dimensionais devem ser utilizadas para suportar variados tipos de inquirições necessárias para o suporte de decisões.

Consequentemente, a chave das tabelas de dimensão deve apresentar o nível de detalhe necessário para dar suporte ao propósito de negócios para o qual o esquema em estrela foi criado. Isso pode ser diferente para múltiplos esquemas em estrela que contenham dados similares, mas em combinações diferentes (William H. Inmon, 1999). Enfim importa tudo aquilo que descrevemos para que o administrador ou gestor do negócio possa trabalhar.

Tabelas de Factos

No DW existe uma tabela de facto que armazena os valores detalhados das medidas (métricas) dos processos de negócio que devem ser analisadas. Ou seja, representa as informações que serão avaliadas, sendo, normalmente, constituída de valores numéricos que representam os objetos da análise. Como por exemplo, uma tabela que armazena Inscrição da disciplina, Avaliação das notas do aluno, entre outras. Na maioria das tabelas de factos domina o esquema em estrela, são nestas tabelas representadas as transações, atividades e acontecimentos. O seu funcionamento encontra-se no centro rodeado de tabelas de dimensões que se ligam através da chave primária. Também composta por dois tipos de atributos que são as chaves estrangeiras e os factos ou dados relativos ao acontecimento específico. Em cada linha desta tabela é apresentada uma medição, sendo que todas as medições de uma tabela terão de ter a mesma granularidade.

Basicamente o nome do modelo dimensional vem a partir das dimensões do negócio (*Business Dimension*) em que precisamos de incorporar o modelo de dados lógico. Esta parte é uma técnica do projeto lógico utilizada para estruturar as dimensões do negócio e as métricas ou factos que são

analisados ao longo destas dimensões. Posteriormente esta técnica de modelagem é intuitiva para o efeito em que também se mostrou para proporcionar alto desempenho das consultas e análises (Kimball, 2008).

O modelo dimensional utiliza o conceito do modelo inter-relacionamento entre entidades com algumas restrições importantes. Cada modelo dimensional composto por uma tabela com uma chave primária composta (tabelas de factos), e um conjunto de tabelas menores (tabelas de dimensões). As características mais importantes neste modelo dimensional são todas as chaves primárias substituídas por uma chave estrangeira (*surrogate keys*). Isto é obviamente ter uma associação entre tabelas de facto com tabelas de dimensão que é dependente a chave estrangeira. A utilidade desta chave estrangeira é permitido ter acesso livremente nos dados de DW ao invés de aquilo que produzido pelo sistema OLTP .

Um sistema OLTP exige a normalização no formato de dados para garantir a consistência dos dados, minimizar o espaço de armazenamento e reduzir as redundâncias, para suportar grandes volumes de transações que se movem muito rápido. O modelo OLTP muitas vezes se parece com uma teia de aranha feita de centenas ou mesmo milhares de tabelas por isso é difícil de entender. Pelo contrário, o modelo dimensional frequentemente usando no DW que utiliza em formato desnormalizado, normalmente utiliza o modelo estrela (*star schema*) e floco de neve (*snow schema*). Esses modelos são fáceis de entender e cumpridos com os requisitos de negócio, suportando consultas simples e proporcionam um desempenho superior nas consultas com a minimização as tabelas de associação (*join*) (Ponniiah, 2002).

O Modelo Estrela (*star schema*)

Como vimos que esse modelo da esquema em estrela é simplesmente um modelo relacional com a relação um para muitos entre a cada tabela de dimensão e a tabela de factos. Esse modelo, a cada tabela de facto conta com todos os níveis de agregação de dados possíveis. As chaves primárias existem apenas para os atributos mais atómicos ou mais elementares. Quando a fazer pesquisa uma tabela então vai possuir todos os atributos associados a essa dimensão. É por isso que se chama modelo em Estrela, onde a tabela de factos fica no centro e cercada por tabelas de dimensão, fazendo lembrar a forma de uma estrela.

A estrutura do esquema de estrela é uma estrutura que pode ser facilmente entendida pelos utilizadores e com os quais podem confortavelmente trabalhar. Normalmente na forma desta estrutura, os utilizadores exibem as suas medidas críticas ao longo das dimensões dos seus negócios (Ponniiah, 2002).

O benefício deste modelo que supera quaisquer deficiências é que o número de operações de “join” entre tabelas é minimizado na generalidade das análises efetuadas. Ou seja, este modelo de estrela (*Star Schema*) é mais simples e de mais fácil navegação. No entanto, este modelo de estrela desperdiça espaço repetindo com as mesmas descrições ao longo de toda a tabela (redundâncias). Pelo contrário, também identifica a desvantagem deste modelo que tem a ver com o facto das tabelas de dimensão possuírem um tamanho muito grande. O aspeto do modelo de estrela ilustra-se como na Figura 5.

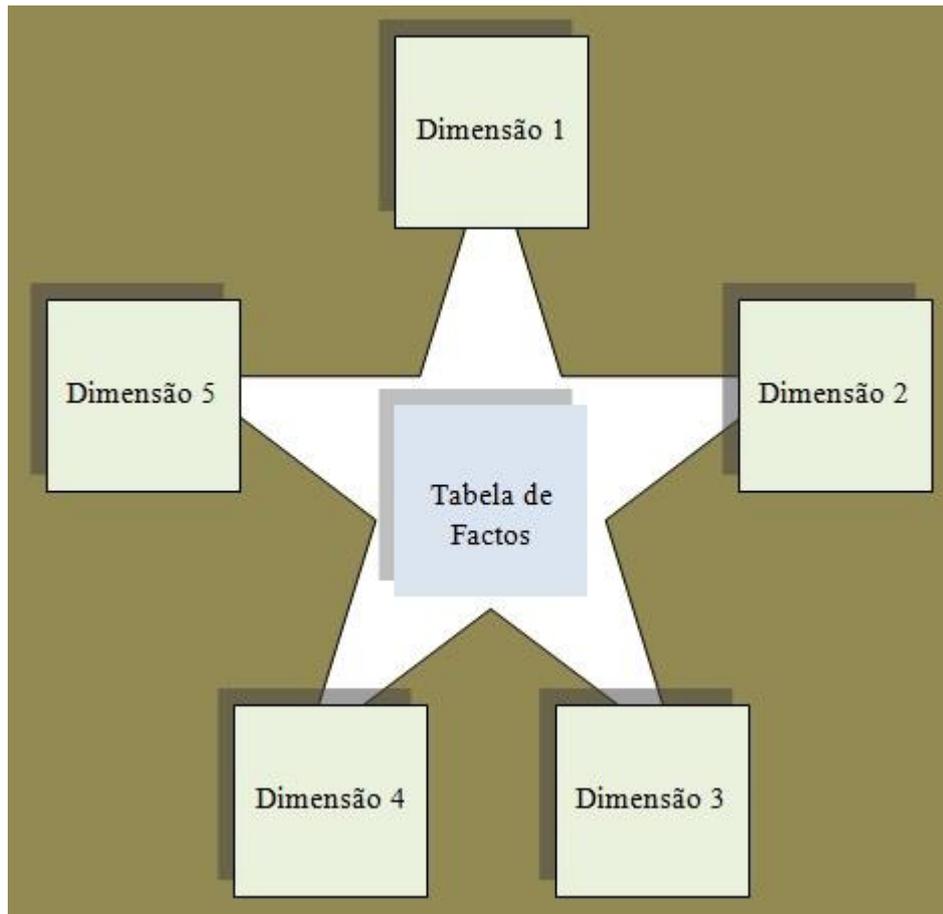


Figura 5 Exemplo Modelo de Estrela (*Star Schema*). Fonte: Elaboração própria

O Modelo Floco de Neve (*Snow Flake*)

Neste modelo, as tabelas dimensionais relacionam-se com a tabela de factos, contudo algumas dimensões relacionam-se apenas entre si, isto ocorre para que ocorra a normalização das tabelas dimensionais, visando diminuir o espaço ocupado por estas tabelas (Raph Kimball, 1998).

A estrutura destes dados é mais complexa do que o esquema em estrela, com a normalização das tabelas de grande dimensão com uma ou mais colunas com dados duplicados. Tabelas de dimensões

são normalizadas para reduzir a redundância de dados, de modo a que a estrutura da tabela se torne mais simplificada. Com este agrupamento, os dados serão mais fáceis de ler ajudando os desenvolvedores de aplicações para criar o desenho da interface do sistema e filtragem de dados.

Esta estrutura vai poupar a capacidade de armazenamento, mas o tempo de execução dos dados demora mais, dado o número de tabelas de dimensão que estão relacionados muito mais e exige adicionamento de relações da chave estrangeira. As consultas tornam-se mais complexas, resultando na diminuição do desempenho. Por tanto, a base de utilizar o DW é a velocidade de acesso e a execução de dados, não é o tamanho de dados que é muito menor ou estrutura de dados que é mais simples. No entanto tem a inteligência na criação da estrutura de dados esquema em estrela ou esquema em floco de neve que irá determinar o desempenho do serviço do DW que eles têm.

O benefício deste modelo tem a ver com o facto das tabelas de atributos serem normalizadas, pequenas e sem redundâncias e diminuído o espaço em disco. No entanto, as análises de níveis mais elevados implicam a realização de um grande número de operações de *join* entre tabelas, o que constitui uma desvantagem. Outra desvantagem é a sua manutenção complexa, dado que cada dimensão se divide em várias outras tabelas organizadas de uma certa forma fazendo lembrar um floco de neve (Ponniah, 2002). O aspeto do modelo de dados floco de neve ilustra-se como na Figura 6.

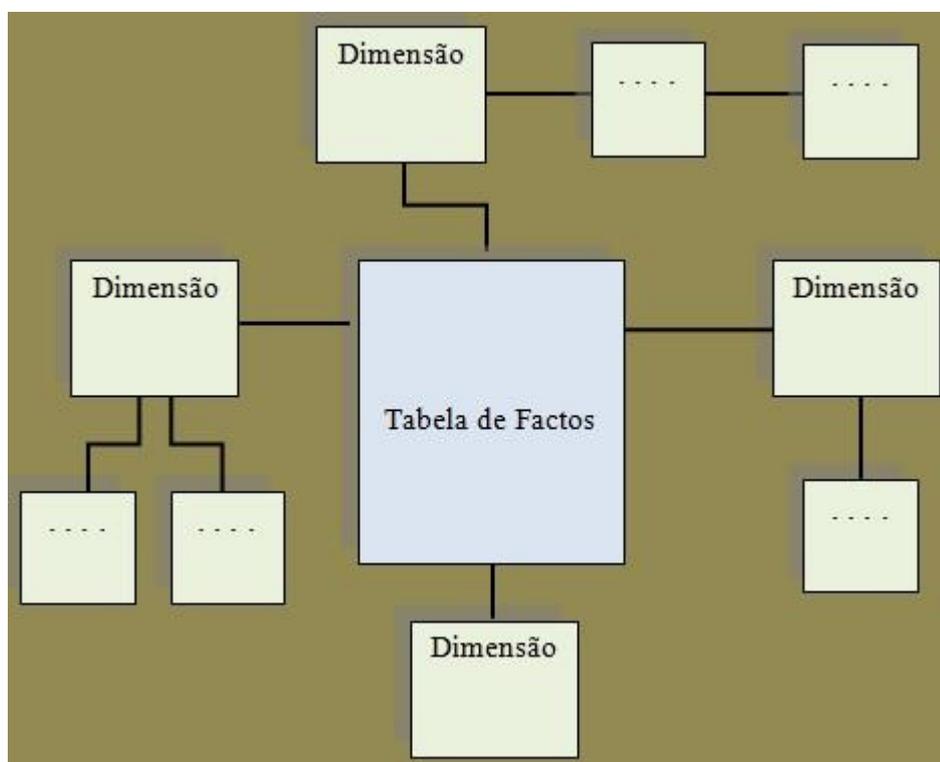


Figura 6 Exemplo Modelo Floco de Neve (*Snow Flake*). Fonte: Elaboração própria

Modelo Normalizado

Neste modelo, apenas os dados elementares ficam guardados em tabelas. Todas as agregações e acumulações de dados não ficam registadas. Se existirem, serão feitas em processos “*on-line*”, em modo “*on-demand*” (ou seja, no pedido do utilizador) (Ralph Kimball, 2002).

2.3.2 Arquitetura de DW

A razão fundamental de construir um DW é promover a qualidade de informação numa organização. Os dados provenientes de fontes internas e externas em vários formatos e estrutura são consolidados e integrados num único repositório. O sistema do DW consiste por DW e todos os componentes utilizados para a construção, avaliação, e manutenção. Para ser útil o DW deve ser capaz de responder a consultas avançadas de maneira rápida, sem deixar de mostrar detalhes relevantes à resposta. Para isso ele deve possuir uma arquitetura que lhe permita recolher, manipular e apresentar os dados de forma eficiente e rápida. Mas construir um DW eficiente, que servirá de suporte a decisões para a empresa, exige mais do que simplesmente descarregar ou copiar os dados dos sistemas atuais para um banco de dados maiores. Deve-se considerar que os dados provenientes de vários sistemas podem conter redundâncias e diferenças, então antes de passá-los para o DW é necessário aplicar filtros sobre eles.

Este estudo de arquitetura Figura 7 permite compreender como o DW faz para armazenar, integrar, comunicar, processar e apresentar os dados que os utilizadores utilizarão em suas decisões. Um DW pode variar na sua arquitetura conforme o tipo de assunto abordado pois as necessidades também variam de instituição para instituição.

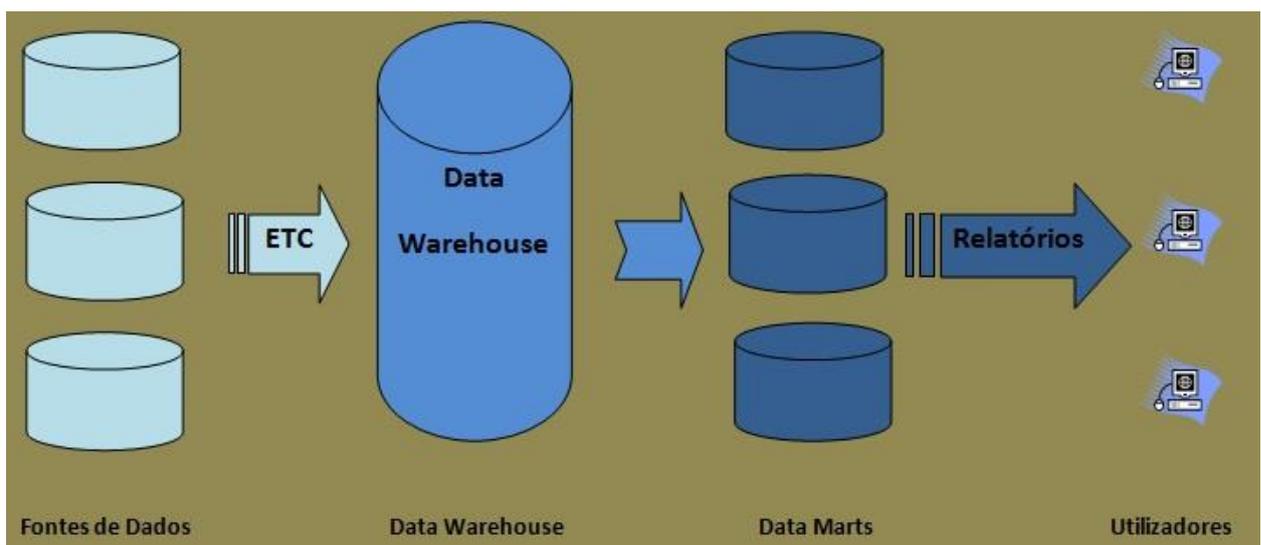


Figura 7 Arquitetura de DW. Fonte: Elaboração própria

Sistemas Operacionais

A primeira fase desta arquitetura é a camada banco de dados operacional que corresponde aos dados das bases de dados operacionais da organização junto com dados provenientes de outras fontes externas que serão tratados para compor o DW. Basicamente as fontes de dados podem ser classificadas como internas ou externas. As que pertencem às internas são aquelas que mais utilizada na organização no seu dia-a-dia, como por exemplo vendas, compras, e as outras. Pelo contrário, os sistemas devem ser considerados como externos ao DW e porque se presume que se tenha pouco ou nenhum controlo sobre o conteúdo e o formato dos dados nesses sistemas. Por exemplo, um simples ficheiro com os códigos postais de Portugal até sofisticados quadros com informação estatística demográfica ou económico-social (Caldeira, 2008). Há várias aplicações que cobrem estas fontes de dados, passando por um simples ficheiro de texto, uma folha de cálculo e até que as diversas bases de dados.

Aquisição de dados (*Data acquisition*)

Esta fase consiste por todos os programas como *Data Transformation Services* (DTS), que são responsáveis pela extração de dados de fontes operacional, preparação e carregá-lo para DW. A fase do componente de acesso consiste por todas as aplicações como OLAP que usa as informações armazenadas no DW.

Já que o DW é usado para tomada de decisão, é importante que os seus dados estejam corretos. A extração de dados é uma das mais demoradas tarefas de desenvolvimento do DW. Dados consolidados a partir de sistemas heterogêneos, entretanto uma vez que grandes volumes de dados estão envolvidos, há uma alta probabilidade de erros e anomalias nos dados. Ainda precisa de ser transformado e limpo antes de carregados para DW. Tamanhos inconsistente de valores, entradas erradas e violação de restrições de integridade são alguns exemplos onde a limpeza de dados torna-se necessária.

A limpeza de dados é uma tarefa essencial no processo do DW, a fim de obter dados corretos e qualitativos no DW. Este processo contém, basicamente as seguintes tarefas que são: Conversão de dados de fontes de dados heterogêneas com várias representações externas a uma estrutura comum adequada para o DW, Identificação e eliminação de dados redundantes ou irrelevantes, Transformando dados para valores corretos, por exemplo, observando-se o uso de parâmetros e consolidando estes valores num formato comum. Conciliar diferenças entre múltiplas fontes, devido

ao uso de homônimos (mesmo nome para coisas diferentes), sinônimos (nomes diferentes para as mesmas coisas) ou unidades de medida diferentes (Kimball, 2004).

Quando o processo de limpeza é concluído, os dados armazenados no warehouse devem ser misturados e colocados a um nível de detalhe comum contendo informações relacionadas ao tempo para permitir o uso de dados históricos. Antes de carregar dados no DW pode ser realizada as tarefas como filtragem, classificação e indexação.

Metadados

A gestão do componente de metadados é responsável para ter acesso de todos os diferentes tipos de metadados. O metadados definiu-se como dados que descrevem o significado dos dados. No DW, há vários tipos de metadados, por exemplo, informação sobre fontes operacional, estrutura e semânticas dos dados de DW, as tarefas executadas durante a construção, manutenção e acesso do DW, etc. Dados sobre as visões dos utilizadores devem poder ter acesso aos dados de um DW sem que tenha que saber onde residem estes dados ou a forma como estão armazenados.

A definição mais comum de metadados que o autor (Inmon, 2002) referiu é os dados sobre dados, ou seja, a descrição da estrutura, conteúdo, chaves, índices e outros dados. Acrescentou-se ainda que os metadados têm alguns fundamentos, incluem: ID de documento, data de entrada no DW, descrição do documento, fonte do documento, a data fonte do documento, classificação de documentos, dados de índice, datas de limpeza, localização física, comprimento do documento e referência relacionadas.

Devido ao grande volume de dados contidos no DW é necessário que exista uma forma flexível e eficiente de acesso aos dados. É necessário saber que dados estão disponíveis e onde estão localizados. Os metadados podem ser vistos como a descrição dos dados, do seu ambiente, como são manipulados e para onde são distribuídos. Os metadados permitem definir as estruturas de informações usadas, os algoritmos utilizados para a transformação, conversão, acumulação e agregação de dados. Permitem identificar as fontes de informação, qual o destino dos dados e também controlar o mapeamento de dados ODS (*Operational Data Store*) para o DW. Os metadados podem servir também para fazer a consolidação dos termos e temas dos dados. Um determinado termo deve ter o mesmo significado em todo o DW.

Armazenamento de dados (*Data storage*)

Devido à natureza especial do DW e de acesso, devem ser adotadas alguns mecanismos para o armazenamento de dados, processamento de consultas, e gestão de. As soluções do DW precisam de

complexos requisitos de consultas e operações envolvendo os grandes volumes de acesso aos dados. Essas operações precisam de métodos de acesso especial, estruturas de armazenamento e técnicas de processamento de consultas.

A implementação de um DW é uma tarefa complexa que contém duas fases principais. O primeiro ponto especificado é a fase de configuração, que é uma visão conceptual do DW de acordo com requisitos do utilizador. Em seguida, as fontes de dados relacionados e o processo (*data acquisition*) *Extraction-Load-Transform* (ETL) são determinados. Por fim, as decisões sobre o armazenamento persistente do DW usando a tecnologia da base de dados e várias maneiras será acedido durante as análises que são feitas (Kimball, 2004).

Depois da carga inicial, a fase de operação, os dados do DW devem ser regularmente atualizados. Isto é, a modificação de dados operacionais durante a última atualização deve ser propagada para o armazém de tal forma que os dados armazenados no DW refletem o estado dos sistemas transacionais.

A camada do DW propriamente corresponde aos dados usados para fins informacionais. O DW é simplesmente uma visão lógica ou virtual dos dados. Num DW que existe fisicamente, cópias dos dados operacionais e externos são armazenadas, de modo a promover fácil acesso e alta flexibilidade de manipulação.

Acesso aos Dados

Basicamente este componente de acesso aos dados ou *front-end*, identifica como a parte final de acesso à informação que possa interagir com o componente da parte operacional através das possibilidades das ferramentas utilizadas. Este sistema é definido como responsável de fornecer soluções de acesso aos dados para sustentar as necessidades dos utilizadores. Atualmente, existem varias ferramentas e produtos que estão disponíveis comercialmente. Em comum todas as ferramentas e a linguagem de cliente utilizada no dia-a-dia para este acesso é o SQL (*Structured Query Language*), Data base Access e Microsoft Excel. As principais aplicações que considerados importantes na execução dos relatórios são ferramentas de OLAP e *Crystal Reports*. No qual tem uma visão dos dados orientados à análise, navegação rápida e flexível que ajudando a interpretar o que ocorreu e decidir sobre estratégias futuras.

Com tudo, a evolução da tecnologia de hardware e software, os sistema de cada dia mais sofisticados para a manipulação de dados, análise de dados e apresentação de dados através das ferramentas de data mining e visualização.

Data Marts

Data Mart é um sistema de suporte a decisão que incorpora um subconjunto de dados em um DW e é um resumo dos dados relativos a um departamento ou uma função específica da organização. O *Data Mart* é geralmente descrito como um subconjunto dos dados contidos em um DW extraído para um ambiente separado (William H. Inmon, 1999).

Logo, um DW é composto por um conjunto de *Data Marts* integrado. *Data Marts* especifica-se nas exigências dos utilizadores em função de departamento ou empresa particular de uma organização. Desde que o *Data Marts* são especializados para operações departamentais, eles contêm dados menor e os utilizadores são muito mais capazes de explorar *Data Marts* do que o DW.

Em termos das empresas, por exemplo, os *Data Marts* têm propósitos específicos relacionados ao negócio, como medida do impacto de promoções de marketing, medida ou previsão de vendas, medida do impacto da introdução de novos produtos e entre outras. De acordo com (William H. Inmon, 1999) dada algumas das razões fundamental de implementar o *Data Mart* ao em vez de o DW são:

- O *Data Mart* permite aos utilizadores analisar os dados de que necessitam na maioria das vezes nas suas operações diárias.
- O *Data Mart* é mais específico e contem menos dados que é bom para o tempo do feedback por parte dos utilizadores muito mais rápido. Assim não tem nada a comparar com o DW em termos de tempo e recursos porque o *Data Mart* é uma tarefa mais simples e mais barata.
- No ponto de vista de Engenharia de Software, considera que a construção de *Data Mart* é um projeto mais viável porque as exigências são muito mais explícitas em vez do projeto de DW com o corporativo mais largo.

Business Intelligence(BI)

BI é definido como a habilidade de uma organização para tirar todas as suas capacidades e convertê-los em conhecimento. Isto produz uma das grandes quantidades de informação que podem levar ao desenvolvimento de novas oportunidades para organização. A quando essas oportunidades foram identificados e uma estratégia tem sido efetivamente implementado, eles podem fornecer uma organização com uma vantagem competitiva no mercado e estabilidade no longo prazo.

As tecnologias de BI fornecem visualizações históricas, atuais e prospetivas das operações de negócios. As funções comuns de tecnologia de *BI* são consultas, processamento de analítico online,

análises, mineração de dados, mineração de processos, processamento de eventos complexos, gestão de desempenho de negócios e *benchmarking* (Oliveira, 2009).

Concepção do Data Warehouse

DW é um armazenamento de dados de suporte à decisão ou multidimensional que projetado para consulta e análise. O DW é otimizado para o processamento de transações que é o domínio de OLTP. O DW é geralmente uma história e análise dos dados gerados a partir de diversas fontes. Neste caso a análise de dados de DW é separada do armazenamento de dados de transações, uma vez que permitindo uma organização para consolidar dados de várias fontes.

Normalmente o DW armazena os dados em muitos meses ou anos para apoiar a análise histórica. Os dados normalmente carregado através da extração, transformação e processamento de carregamento (ETL) de uma ou mais fontes de dados como aplicações OLTP, aplicações de mainframe ou provedores de dados externos.

Os utilizadores do DW normalmente analisam os dados de forma frequentemente associados como por exemplo incluem a consolidação dos números do ano passado. A análise mais sofisticada é incluindo análise de tendência e mineração de dados, que utiliza dados existentes para estimar as tendências ou prever o futuro. Esse ambiente de DW geralmente fornece uma base para a inteligência de negócios. Basicamente tem algumas principais características de DW que avalia o seu desempenho mais rápido da consulta com a transferência de dados elevada, são: a maioria dos dados em de normalização para simplificar e melhorar o desempenho, utilizar grandes quantidades de dados para dados históricos e query é executada para recuperar os dados.

Após este processo, os dados consolidados podem ser importados para o DW usando um dos carregadores de dados em massa, um aplicativo personalizado ou um assistente de importação / exportação fornecidos pelas aplicações de administração de DBMS. A carga do DW tem que lidar com volumes de dados muito maiores que os bancos de dados operacionais. A fim de sustentar os carregamentos, faz o *refresh* do DW que consiste em propagar as atualizações ocorridas nos bancos de dados operacionais para o banco de dados derivado do DW.

2.3.3 Extração Transformação e Carregamento

O ambiente de DW é muito importante pelo tratamento de dados com outras áreas da tecnologia e responsáveis pelo desenvolvimento dos sistemas. Essencialmente o conteúdo principal de um DW é contém dados sumarizados, históricos e detalhados para suportar a tomada de decisões táticas e

estratégicas. Ainda se lembram que no ambiente das funções e serviços da arquitetura técnica do DW, dividimos em três áreas funcionais que são aquisição de dados, o armazenamento de dados e entrega de informações.

Extração de dados, transformação de dados e carregar os dados englobam as áreas de aquisição e armazenamento de dados. Neste caso, os dados são inicialmente extraídos de sistemas operacionais e de fontes externas, posteriormente integrados e transformados antes de serem carregados no DW. Normalmente os dados recebidos em formato texto (ASCII) que são filtrados para a identificação de diversos erros, ou seja, incluem todas as funções e procedimentos para alterar a fonte de dados para os formatos exatos e estruturada antes dos seus continuções dos processos (Kimball, 2004).

No processo das transformações de dados obtêm-se uma fase de limpeza, eliminação, combinação, validação, consolidação, agregação e sumarização. Após a transformação de dados, estes processos são compostos por todas as funções para mover fisicamente os dados ao repositório de DW. Finalmente, os utilizadores acessam o DW através de ferramentas de *front-end* ou aplicações submetendo suas consultas, de modo a obterem informações que permitam a tomada de decisões.

A Figura 8 demonstra-se o ambiente DW do processo ETC até a ter acesso aos utilizadores.

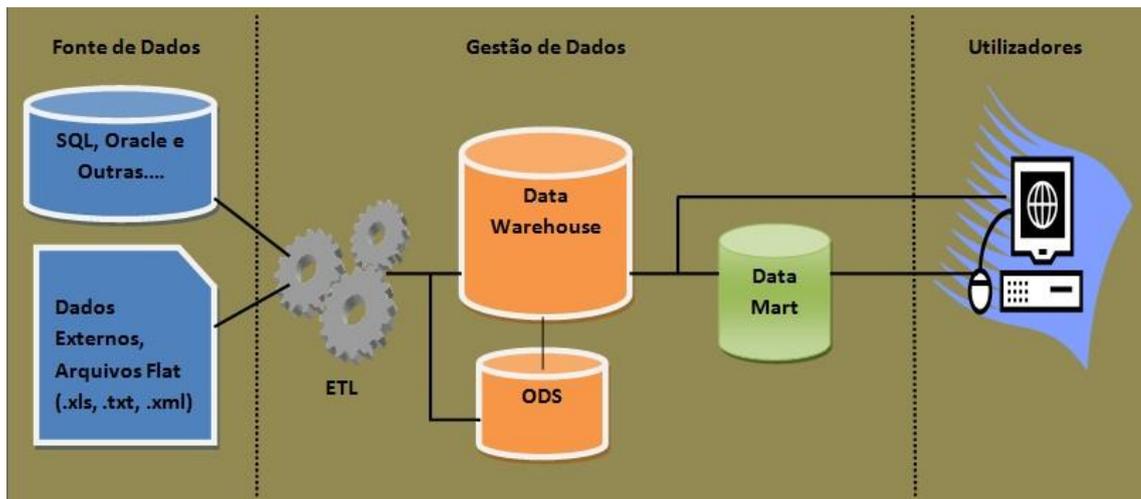


Figura 8 Extração, Transformação e Carregamento de Dados (ETC). Fonte: Elaboração própria

Extração

Este processo de extração é a fase inicial aos gestores na obtenção de dados para o desenvolvimento do DW. O sentido de extrair é ler e compreender fontes de informação e copiar as partes necessárias para a área de transformação de dados, a fim de serem trabalhadas posteriormente. É comum que os dados extraídos não façam de referência a realidade, uma vez que extrações para o

DW não costumam ter a devida atenção dos desenvolvedores dos sistemas transacionais (Kimball, 2004).

Essas extrações podem demandar operações complexas, e alteração no formato dos dados, ou até mesmo a criação de informações novas a partir de operações ou cálculos específicos. Os dados provêm de várias fontes e independentes que normalmente podendo ser as fontes de base de dados dos sistemas transacionais, folha de Excel, ficheiro de texto e entre outras. O que requer é a sua tradução e adequação para a estrutura padrão do armazenamento.

Muitas vezes na extração de dados enfrenta um grande desafio é que determinar quais são os dados para extrair e que tipo de filtros que vai implementar. Por fim, considera que é uma das atividades que consomem mais tempo no desenvolvimento do DW. Existem algumas maneiras de extração, entre elas, são através execução de códigos do programa construídos sobre um sistema fonte de modo a gerar arquivos com os dados desejados e por outra maneira é utilizar uma ferramenta de extração específicas com códigos próprios para execução sobre o sistema fonte de dados de forma a obter dados necessários.

Transformação

Após as extrações os dados admitem transformações internas no DW. A transformação de dados é efetuada com base em regras de negócio definidas com o objetivo de obter a informações necessária para o negócio. Dentre as transformação de dados em bruto envolve a adaptação, limpeza e consolidação da informação antes de ser integrada no DW (Kimball, 2004). Neste caso existe também a unificação de padrões como por exemplo identificar M para masculino e F para feminino. Essa integração de dados conforme regras definidas pelos analistas de negócios em conjunto com os desenvolvedores. O objetivo principal nesta fase de transformação é eliminar as diferenças semânticas entre os dados extraídos e o esquema multidimensional adotado. Essa transformação poderá envolver um ou diversos processos dependem as necessidades que na maioria são a limpeza, eliminação, combinação, desnormalização e normalização, cálculos, derivação e alocação (Daniele Del Bianco Hokama, 2004).

Na limpeza de dados obviamente é um conjunto de atividades deve ser realizada para corrigir o uso incorretos de códigos e caracteres especiais, tratar dados perdidos e corrigir os valores duplicados ou errados. Esse tipo de limpeza é para manter os dados no formato padronizado, não duplicados, corretos e consistente. Por exemplo, o nome de uma cidade que é incompatível com o seu código postal ou sobreposições, ou mesmo dois campos representado a mesma informação com formatos diferentes.

A questão de eliminação dos dados é constitui por desconsiderar os campos e dados desnecessários ao sistema de DW. Uma vez que o sistema de DW precisa mesmo os dados que são úteis e definidas para fins das consultas. A realizar também combinação da informação é que quando as fontes de dados possuem exatamente com os mesmos valores de chaves representado registos iguais ou complementares. Existe entretanto o padrão de normalização neste processo de transformação com objetivo de reunir as estruturas de dados que separadas em diversas tabelas de dimensões. Em fim nos cálculos, derivação e alocação aplicam as transformações num conjunto de funções das regras dos negócios identificadas durante o processo de levantamento de requisitos. Por exemplo a fazer manipulação de textos, aritmética de data e hora, entre outras (Daniele Del Bianco Hokama, 2004).

Carregamento

Por fim, o processo de carregamento no DW é realizado após efetuados todos os tratamentos de dados nos processos de extração e transformação. A parte de carregamento dos dados também possui uma enorme complexidade que deve ser considerada aos fatores importantes como a integridade dos dados, os tipos de carregamento a ser realizada, otimização e o suporte completo deste processo (Kimball, 2004). O primeiro fator é fundamental para rever as chaves estrangeira dos dados existentes de acordo com a chave primária das respectivas tabelas. Esse tipo de processo de integridade facilita aos desenvolvedores controlar de maneira mais eficaz a redundância dos dados. Outro fator que deve saber é o tipo de carregamento a ser realizada, que são do tipo incremental ou carregamento total. Estes dois tipos têm as diferenças que o carregamento incremental normalmente é feito por tabelas de factos e outra é feita em tabelas de dimensão. O carregamento total é depende da necessidade do negócio em questão porque o analista terá que excluir os dados existentes e depois inclui-los novamente .

Uma das mais valia é otimizar o processo de carregamento que possui por um conjunto de técnicas para toda a base de dados, nomeadamente a evitar geração de log durante o processo, criar índices e agregar dados. Essa característica é feita através da utilização de ferramenta sobre a área de organização de dados. Por último o serviço de carga também precisa suportar as exigências que ocorreram antes e depois da carga atual, como por exemplo eliminar e recriar índices e particionamento físico de tabelas e índices (Kimball, 2004).

Capítulo 3 Ferramentas Utilizadas Para O Caso Prático

Para a construção do estudo de caso, foram selecionadas algumas ferramentas dentre as disponíveis. A seleção não levou em conta quais as melhores ferramentas ou as que possuíam melhor desempenho. Esta seleção levou em consideração o conhecimento das pessoas envolvidas no estudo de caso. Todo processo de construção de DW foi efetuado através do recurso de ferramenta *BI* da Microsoft, o SQL Server 2012 e o Microsoft Office 2007, que contêm todas as funcionalidades necessárias à construção da plataforma analítica OLAP.

O *Business Intelligence Development Studio* (BIDS) é uma ferramenta utilizada para desenhar, construir e testar o pacote de dados. Este *BIDS* fornece um ambiente de desenvolvimento orientado ao gráfico que permite a criação, implementação, análise e construção de relatórios do DW. Como por exemplo as potencialidades específicas desta tecnologia do Microsoft constituídas por SQL Server Integration, Reporting Services e Analysis Services.

3.1 Microsoft SQL Server

Essa é a principal ferramenta utilizada neste trabalho para criar a base de dados do modelo relacional. Sem dúvida que essa ferramenta é melhor, dada a maior disponibilidade e facilidade no uso. Sabendo que a maioria dos programadores utiliza essa ferramenta de SQL Server 2012 com uma potencialidade enorme e possui as características necessárias para um SGBDR. Existem algumas das suas características que composto por, a alta disponibilidade e proteção dos dados, desempenho incrível e rápido, exploração e visualização rápida dos dados e dados consistentes (Magalhães, 2013). Geralmente, essa ferramenta traz maior desempenho, disponibilidade, segurança e inteligência aos utilizadores. Por outro lado, o conhecimento suficiente do autor é a maior vantagem na escolha.

3.2 SQL Server Analysis Services (SSAS)

O *BI* começa do processo de acumulação dos dados (ETC) seguidos pela apresentação dos dados e, em seguida, a interpretação dos dados na forma de análise. Claro que, tudo isso deve ser suportado com os dados certos, no momento certo e com formato certo. *BI* é uma ferramenta eficaz para apoiar as decisões de negócios, seja na forma de decisões estratégicas, táticas ou operacional. Analysis Services é uma tecnologia para OLAP e mineração de dados. O processo de administração OLAP realizado no SQL Server Management Studio de forma Viewing data, fazer multidimensional Expression (MDX), Data Mining Extension (DMX) e XML for Analysis (XML/A) e definir o regulamento ao acesso de segurança de dados OLAP. Assim, para o desenvolvimento de OLAP

utiliza-se o *BIDS*. Com a utilização de *BIDS* podem criar componente core de Analysis Services (Data Source, Data Views, Dimension, Value, Cube, Role e entre outros) (Magalhães, 2013).

3.3 Microsoft Office Excel

Em geral, o Microsoft Excel é um programa de folha de cálculo que foi produzido pela Microsoft que, hoje em dia, considerado como um dos mais populares aplicativos de computador. A versão utilizada foi de 2007 do SP1. Os seus recursos incluem uma interface intuitiva e capacitadas por ferramentas de cálculo e desenho gráficos. Este aplicativo é mais utilizado no trabalho prático da dissertação como recurso para fornecer os dados de origens e transformar os dados da folha de cálculo para SQL Server através da linguagem SQL. Baseia-se nos dados recolhidos da UNTL, atualmente, todos eles utilizam a ferramenta de Excel para registar, visualizar e arquivar os dados dos estudantes, professores e entre outros. Portanto, com essa funcionalidade permite interagir ao DW de forma direta e torna-se uma solução deste trabalho.

3.4 Ferramentas Front End

O componente *front end* de um sistema de DW é o responsável por fornecer uma solução de acesso aos dados por parte dos utilizadores que atenda as suas necessidades. As ferramentas *front end* são utilizadas para análise, ajudando a interpretar o que ocorreu e a decidir sobre estratégias futuras. Neste tipo de aplicação, somente a operação de consulta se faz necessária que a profundamente selecionar o conjunto de dados necessários, fazer cálculos e manipulação dos dados e apresentação das informações.

Atualmente, a ferramenta OLAP é mais bem conhecida no desenvolvimento das aplicações para o sistema de DW. Saiba que, com a evolução da tecnologia essa ferramenta ficou mais considerado como o recurso útil no ciclo da vida dos programadores.

3.5 Microsoft SQL Server Report Builder

O Report Builder 3.0 oferece uma série de melhor forma no desenho de relatórios. Essa ferramenta tem o próprio *framework* que é apropriada e bem interessante, por exemplo, pode adicionar mapas, criar gráficos, girar o texto em 270 graus, quebrar a página de controlo e compartilhar conjuntos de dados de relatórios guardados no servidor para outros utilizadores. Portanto, ao longo deste projeto foi feito utilizar o Report Builder, como uma solução simples na criação de todos os relatórios.

Capítulo 4 Estudo de Caso Prático

4.1 Problema

Hoje em dia, com a evolução da tecnologia as empresas e as instituições promovem e modernizam os seus sistemas de informações a um ponto mais convenientes. Mesmo ainda existem alguns que se utilizam os seus mínimos recursos para obter informações necessários. A olhar ao sistema de DW deste projeto, exige a qualidade de dados ou informações que é considerada como um dos sucessos para suportar à tomada de decisão na instituição em questão.

Considera-se neste trabalho, realizará um estudo de caso no instituto do ensino superior da UNTL e levaria alguns fatores importantes que incentivam o autor. Primeiro, a criação do sistema de DW é considerado como uma prioridade para esta instituição a ganhar a sua qualidade de informação. Segundo, facilitar aos diretores competentes nesta instituição à tomada de decisão de forma eficiente e eficaz. Sendo assim, na análise de requisitos feita à UNTL, verificou-se muitos dados em falta e não estavam em condições para a criação do sistema de DW.

Por outro lado, a existência de todas as bases de dados operacionais são registados em folha de cálculo Excel. Esta é um dos problemas que devem ser tratados neste tipo de projeto de DW. Portanto, o autor elabora uma ideia de criar uma base de dados relacional no SQL Server, que baseada nas informações disponíveis nesta instituição, como se ver no anexo 2.

Esses dados do Excel, prejudica a inserção direta dos atributos nas tabelas criadas. Com esta questão, define-se uma maneira de tratar é converter esse tipo de dados para o tipo de dados criados no SQL Server através da linguagem SQL. Tendo sido este o ponto de partida para o estudo da modelação a implementar. Para a modelação de qualquer sistema é fundamental estudar a sua realidade, tal como foi previamente feito.

4.2 Base de Dados

A questão de base de dados é fundamental para o desenvolvimento do processo de DW. Esses dados são bases dos sistemas de origem onde irá utilizado para dar carga nas tabelas de factos e tabelas de dimensão. Sabemos que, a área de trabalho da base de dados e do DW podem ter arquiteturas próprias, de forma a melhorar o desempenho das consultas. Como indica na sessão 4.1 Problema, um dos problemas deste estudo de caso foi a obtenção dos dados. De facto, todas as fontes de dados são do formato Excel é que, prejudicam o carregamento direto nas tabelas criadas em

SQL Server. Inicialmente na figura 9, apresentou-se um exemplo de organizar os dados de fonte, por exemplo, separar a coluna “*FATIN NO DATA MORIS*”(em português significa “Lugar/Local e Data de Nascimento”) para uma coluna de cada que é, Lugar de Nascimento e Data de Nascimento. Essa organização de dados foi feita em Linguagem Java, veja os códigos no anexo 49.

UNIVERSIDADE NACIONAL TIMOR LOROSA'E						
NARAN NO NÚMERU REJISTU ESTUDANTE (NRE)						
SEMESTRE KLIK (I) T.A. 2012						
FAKULDADE DE ENGENHARIA, CIENCIAS E TECNOLOGIA						
DEP. TEKNIKA INFORMATICA						
NO	NRE	NARAN	SEXO	FATIN NO DATA MORIS	DISTRITO	OBSERVA SAUN
1	08.b.04.001	Abilio Godinho	F	25-04-1988	Bobonaro	
2	08.b.04.002	Adelaide Maria Filipe da Cunha	F	Oe-Cusse, 25/8/1989	Oe-cusse	
3	08.b.04.003	Adolfina de Jesus	F	Tutuala, 1/3/1987	Lautem	
4	08.b.04.004	Afonso Ximenes Marcal	M	Liho, 18/10/1989	Ermera	
5	08.b.04.005	Aires Filomeno S. B. de Araujo				La rejistu x 6
6	08.b.04.006	Alda M. da Costa Belo Faria	F	Uai-Hare-Boo, 19/6/1986	Baucau	
7	08.b.04.007	Alexandre Jose Varela	M	Venilale, 9/3/1988	Baucau	
8	08.b.04.008	Amos Andinho Martins	M	Same, 27/8/1987	Manufahi	
9	08.b.04.009	Ana Paula dos Santos Soares	F	Açumano, 25/11/1989	Liquica	
10	08.b.04.010	Andre Anuno Nono				La rejistu x 6
11	08.b.04.011	Angelo Venseslau	M	Uatolari, 21/5/1985	Viqueque	
12	08.b.04.012	Antonieta F. R. Soares	F	Dili, 17/2/1988	Dili	
13	08.b.04.013	Antoninho Lemos da Cunha	M	Manatuto, 8/5/1988	Manatuto	
14	08.b.04.014	Benjamin de Jesus Mau	M	Maliana, 30/3/1989	Bobonaro	
15	08.b.04.015	Carlos Amandio Belo	M	Uatolari, 28/2/1985	Baucau	

Figura 9 Dados dos Alunos de Informática em Excel

Após de organizar esses dados, é continuará com a conversão de dados para que possa carregar na base de dados criado em SQL Server. Os dados de pesquisa não foram suficientes, com muito menos atributos e entidades como referiu na sessão

4.1 Problema, o autor aumentou algumas entidades e atributos como na Figura 10 para necessitar o estudo de caso prático deste projeto.

= "insert into Aluno

(NumAluno, Nome, DataNasc, NumBI, Morada, Localidade, Distrito, Naturalidade, Sexo, NumTelemovel, Email, Estado, AnoEntrada, AnoSaida, Observacoes) values

("&""&B2&""&","&""&C2&""&","&""&F2&""&","&""&"Null"&""&","&""&"Null"&""&","&""&"Null"&""&","&""&G2&""&","&""&E2&""&","&""&D2&""&","&""&"Null"&""&","&""&"Null"&""&","&""&I2&""&","&""&H2&""&","&""&"Null"&""&","&""&J2&""&");"

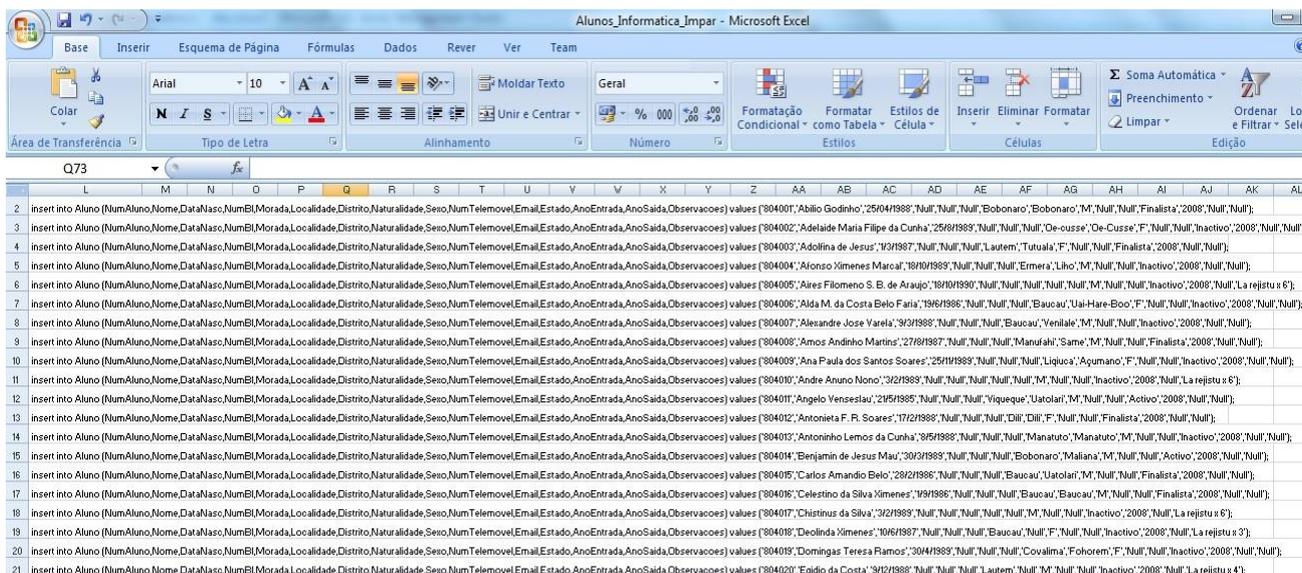


Figura 10 Conversão de Dados com Linguagem SQL

No entanto, o processo de transformação dos dados foi feito através das instruções INSERT INTO do SQL. Esses códigos foram criados numa coluna vazia do Excel e por fim copiar o resultado da execução Figura 10 para uma nova consulta em SQL Server 2012 Figura 11.

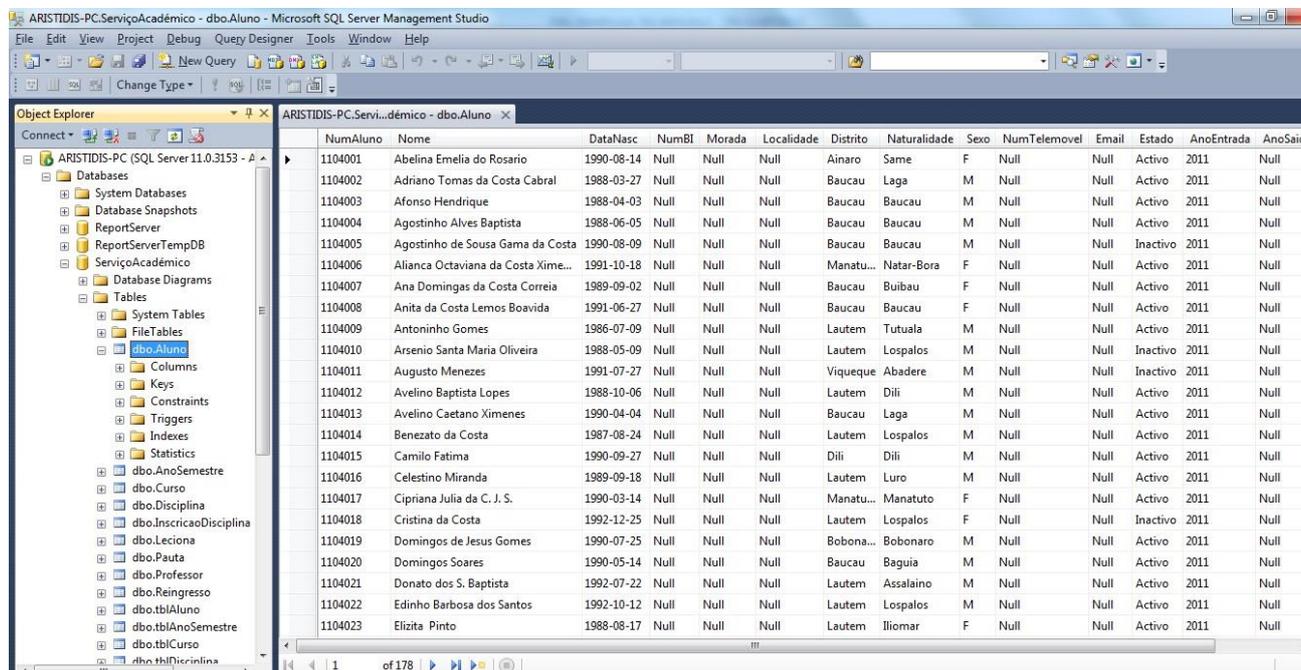


Figura 11 Exemplo Dados dos Alunos no SQL Server 2012

Ao longo deste trabalho, foi construído uma pequena base de dados relacional para instituição da UNTL, principalmente para o curso de Engenharia Informática. Essa base de dados consiste por várias entidades entre elas, são Aluno, Professor, Curso, Disciplina, Pauta, Reingresso, Leciona, AnoSemestre e InscriçãoDisciplina. A cada entidade consiste por vários atributos que interligam uns aos outros através da sua chave primária, veja na Figura 12. Com este modelo relacional é ideal para construir o sistema de DW que foi feito neste projeto.

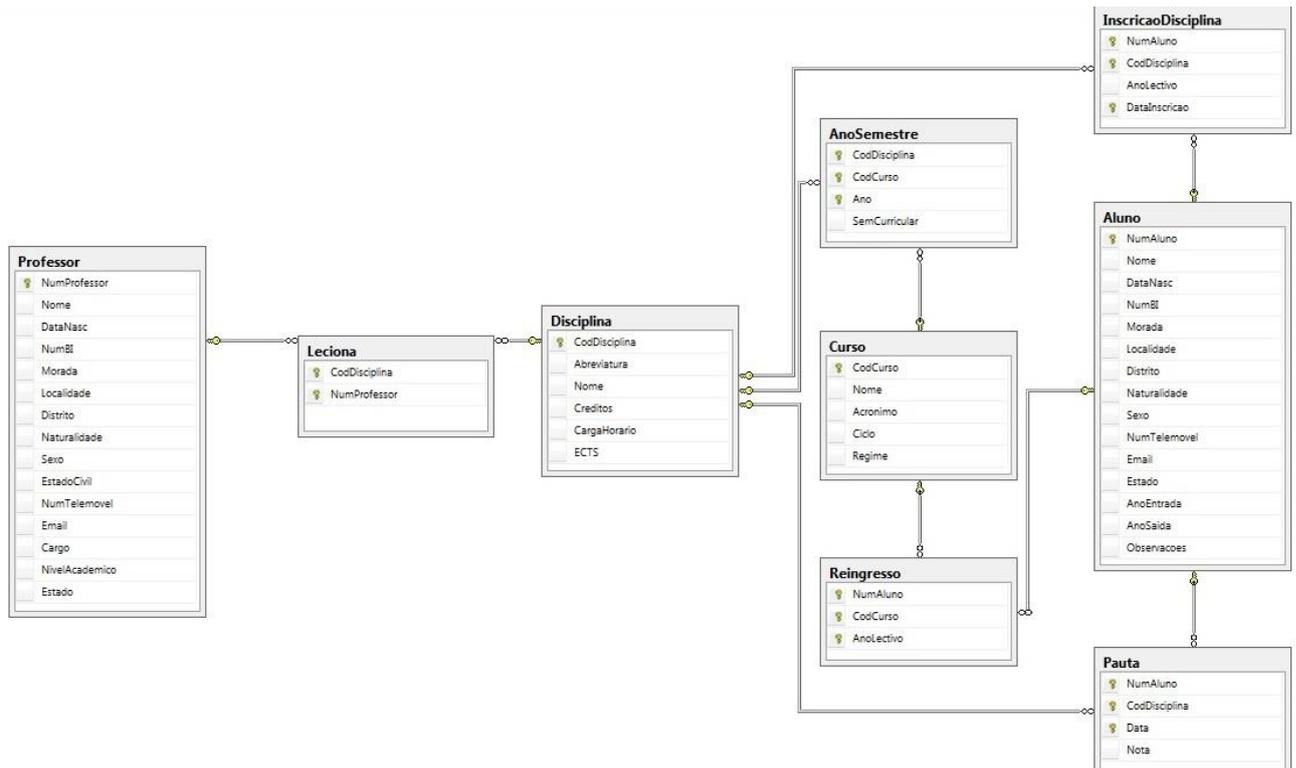


Figura 12 Diagrama de Base de Dados Relacional

4.3 Análise do Processo de Negócio

Processo de Negócio

Esta fase do processo de negócio ou atividade é necessária para quem elaborar o DW de maneira a identificar, planear e estruturar as ideias e determinar concretamente as prioridades a fim de evitar perda de tempo, repetição de atividades ou tarefas, entre outras. Por outro lado, define também o processo de negócio que é uma atividade que se desenrola numa organização com o propósito de atingir um ou mais objetivos previamente definidos (Caldeira, 2008). E portanto, essa atividade previamente estabelecida cujo objetivo é determinar como é que o trabalho será realizado numa organização.

Matriz em Bus

Para elaborar facilmente um DW precisa-se da matriz em Bus, como um suporte informativo onde se identificam os processos chave ou eventos existentes numa determinada organização e se cruzam com propriedades conectados ao seu funcionamento de dimensões de forma a identificar pontos comuns e assim possibilitar a construção de *Data Marts*. A matriz em bus revela-se assim como uma representação gráfica do funcionamento da organização pois decompõe a organização em processos chave e relaciona-os com propriedades ou agentes intervenientes (Sousa, 2010). Basicamente, os processos podem estar relacionados com várias dimensões que por sua vez podem ser comuns a vários processos.

Por outro lado, a ideia dos autores (Caldeira, 2008) e (Ralph Kimball, 2002); referiram-se a matriz em Bus como uma ferramenta que é utilizada para criar, documentar e comunicar a arquitetura do DW onde as linhas de matriz em bus identificado como processo de negócio e as colunas representam as entidades dimensional. O cruzamento dessas dimensões que são relevantes para o seu processo de negócio está marcado um sinal “X” para indicar a relação entre as dimensões de um determinado processo de negócio. A matriz é uma ferramenta que ajuda os utilizadores em processos de planeamento e comunicação. Embora a matriz consista em linhas e colunas, a matriz era capaz de definir a arquitetura geral dos dados para o DW.

A questão de matriz considerado como um plano geral que é projetado de uma forma compacta que pode ser usado para ajudar os seniores de IT e gestores de negócios comunicar o projeto. Na criação de matriz em bus no DW obviamente tem recurso híbrida (*Hybrid Resource*) que é definido por parte dos meios de desenho técnicos, parte de ferramentas para gerir projetos e por meios de comunicação (Ralph Kimball, 2002).

A definição profunda de matriz em bus é uma ferramenta de representação que deve acompanhar o DW, não apenas durante o seu desenvolvimento mas durante todo o seu prazo de vida útil e deve refletir sempre o seu atual estado pois permite uma visão rápida sobre a atual arquitetura. Na Tabela 2, apresenta uma matriz para o curso de Engenharia Informática na UNTL. Essa matriz é composta por linhas que correspondem aos processos de negócios realizados e as colunas correspondem as dimensões ou entidades que participam esse processo.

Tabela 2 Matriz em BUS

Processo de Negócio	Dimensões									
	Aluno	Professor	Disciplina	Curso	Pauta	Ano Semestre	Inscrição Disciplina	Leciona	Reingresso	Total
Avaliação Aluno	X		X	X	X	X			X	6

Pauta										
Inscrição na Disciplina	X	X	X	X			X			5

4.4 Camada de Visualização de Dados

Esse componente da tecnologia desempenha um papel fundamental no processo do DW. E também costuma ser chamada por *Front-end* ou é a interface simplificada que apresenta os dados disponíveis ao utilizador, que não exige mais que alguns cliques para a execução de relatórios. Pensando nesta questão, em décadas passadas, houve o desenvolvimento de uma linguagem própria para a extração de informação que se chama Linguagem SQL. Hoje em dia, com a evolução da tecnologia, essa linguagem passou a ser complexa demais para que a alta direção de empresa ou organização pudesse ter acesso.

Para a construção de DW a infraestrutura que se aplicou foi a referida no Capítulo 3, tendo sido à plataforma de BI de construção relatórios, o Report Builder 3.0, onde o utilizador final terá à disposição o manuseamento de relatórios já existentes, ou a permissão de através destas, construir novos relatórios por ele parametrizados. De facto, a camada de visualização de dados OLAP para este estudo de caso foi destacado na Figura 15, Figura 16 e Figura 17.

4.5 Desenvolvimento do Modelo Dimensional

Após de definir o processo de negócio já agora continuar para a fase de construção do modelo dimensional. Esta é uma das mais-valias no sistema de DW que normalmente consome mais tempo a implementar. Portanto, o esquema utilizado neste modelo dimensional é o esquema de estrela (*star schema*), onde existe uma tabela de factos e rodeada por várias tabelas de dimensão. O motivo por ter escolhido é que a consulta seja mais fácil de navegar aos dados da tabela dimensional e além disso não requer o sub-dimensão.

De seguida irá abordar de forma mais detalhada as descrições dos dados e as suas respetivas dimensões e tabelas de factos que compõem o modelo multidimensional criado para atender a solução proposta. Cada dimensão é definida pela sua única chave primária, que serve como base para manter a integridade referencial quando se liga à tabela de factos.

Dimensão Aluno

A dimensão de aluno contém a informação referente aos alunos do curso de Engenharia Informática da UNTL. Esta dimensão possui por diversos processos de negócios e que identifica

como um dos principais tabela no desenvolvimento de DW da Universidade. Na seguinte Tabela 3 apresentamos um vasto conjunto de atributos da dimensão aluno.

Tabela 3 Descrição de Dimensão Aluno

Nome do Atributo	Descrição	Exemplo
NumAluno (varchar)	Número que identifica o aluno	804001
Nome (varchar)	Nome completo do aluno	Abilio Godinho
DataNasc (date)	Data de nascimento do aluno	25/04/1988
NumBI (varchar)	Número do bilhete de identidade do aluno	121000074849
Morada (varchar)	Morada do aluno	Kampung Baru
Localidade (varchar)	Localidade do aluno	Comoro
Distrito (varchar)	Distrito de nascimento	Bobonaro
Sexo (char)	Sexo do aluno	M
Naturalidade (varchar)	Naturalidade no BI do aluno	Bobonaro
NumTelemovel (varchar)	Contacto telefónico móvel do aluno	75120849
Email (varchar)	Conta e-mail do aluno	abilio.godinho@gmail.com
AnoEntrada (varchar)	Ano em que o aluno ingressou na UNTL	2008
AnoSaida (varchar)	Ano em que o aluno saiu da UNTL	
Estado (varchar)	Estado do aluno	Finalista
Observacoes (varchar)	Descreve a situação do aluno, se está inscrito ou não na universidade	Está inscrito

Dimensão Curso

A dimensão do curso é útil na formação dos processos de negócios. Na Tabela 4 apresenta as descrições dos atributos dos respetivos cursos que existem na Faculdade da Engenharia.

Tabela 4 Descrição de Dimensão Curso

Nome do Atributo	Descrição	Exemplo
CodCurso (varchar)	Código utilizando para identificar o curso	03EI
Nome (varchar)	Nome do curso	Engenharia de Informática
Acrónimo (varchar)	Designação de dois caracteres que permite	EI

	identificar o curso	
Ciclo (varchar)	Designação do ciclo de estudos a que o curso pertence	1º Licenciatura
Regime (varchar)	O funcionamento do curso que pode ser diurno e noturno	Diurno

Dimensão Disciplina

A unidade curricular também é muito importante na avaliação do desempenho e da qualidade de uma instituição superior. Desta forma, na maioria da instituição utilizada os padrões necessários tanto nacional ou internacional que quase definidas pelo concelho científico. Veja na Tabela 5 apresenta as descrições dos atributos de dados que representam as disciplinas dos cursos ministradas na Engenharia Informática da UNTL.

Tabela 5 Descrição de Dimensão Disciplina

Nome do Atributo	Descrição	Exemplo
CodDisciplina (varchar)	Código utilizando para identificar a disciplina	EI24101
Abreviatura (varchar)	Abreviatura definida para a disciplina	SO
Nome (varchar)	Nome da unidade curricular no plano de estudos	Sistemas Operativos
Creditos (varchar)	Número de créditos atribuídos pela obtenção de nota positiva na disciplina	5
CargaHorario (varchar)	A carga de hora para cada disciplina durante um semestre	60 horas
ECTS (varchar)	Número de créditos a cada disciplina	5 ects

Dimensão Professor

A Tabela 6 da dimensão Professor armazena a informação referente aos professores da Engenharia Informática da UNTL e que valia ao processo de negócio do sistema de DW. Neste sentido, a sua função é responsável da unidade curricular.

Tabela 6 Descrição de Dimensão Professor

Nome do Atributo	Descrição	Exemplo
NumProfessor (varchar)	Número que identifica o professor	1111-5
Nome (varchar)	Nome completo do	Marcelino Caetano

	professor	Noronha, M.cs
DataNasc (date)	Data de nascimento do professor	15/05/1976
NumBI (varchar)	Número de bilhete de identidade do professor	121000045740
Morada (varchar)	Morada do professor	Bairopite
Localidade (varchar)	Localidade do professor	Bairopite
Distrito (varchar)	Distrito em que o professor pertence	Lautem
Naturalidade (varchar)	Naturalidade no BI do professor	Luro
Sexo (varchar)	Sexo do professor	M
EstadoCivil (varchar)	Estado civil do professor	Casado
NumTelemovel (varchar)	Número de contacto do professor	77658993
Email (varchar)	Contacto de email do professor	Marcel_pmis@yahoo.com
Cargo (varchar)	Cargo atual do professor	Director
NivelAcademico (varchar)	Grau académico que o professor obteve	Mestrado
Estado (varchar)	Estado atual da profissão do professor com a universidade	Permanente

Dimensão Pauta

A tabela dimensão Pauta armazena as informações necessárias sobre a pauta final de cada aluno. Esta dimensão composta por vários atributos, onde os mais relevantes constam da Tabela 7.

Tabela 7 Descrição da Dimensão Pauta

Nome do Atributo	Descrição	Exemplo
NumAluno (varchar)	Número que identifica o aluno	804001
CodDisciplina (varchar)	Identifica o código da disciplina	EI24101
Data (date)	A data de lançamento de notas	16/7/2008
Nota (varchar)	A classificação final da disciplina de um aluno	14 valores

Dimensão AnoSemestre

Esta dimensão de AnoSemestre na Tabela 8, consiste por atributos importantes como Ano e Semestre curricular e que também correspondem com as dimensões do Curso e Disciplina.

Tabela 8 Descrição da Dimensão AnoLectivo

Nome do Atributo	Descrição	Exemplo
------------------	-----------	---------

CodDisciplina (varchar)	Código que identifica a cada disciplina	EI24101
CodCurso (varchar)	Código que identifica a cada curso	03EI
Ano (varchar)	Ano em que o aluno está a frequentar essa disciplina	1º Ano
SemCurricular (varchar)	O semestre corresponde ao ano em que o aluno está a frequentar	1º Semestre

Dimensão Reingresso

Esta dimensão Reingresso na Tabela 9 é útil, de modo que, facilitar a avaliação dos alunos através dos seus dados ao nível de ano letivo em que fez reingresso.

Tabela 9 Descrição da Dimensão Reingresso

Nome do Atributo	Descrição	Exemplo
NumAluno (varchar)	Número que identifica o aluno	804001
CodCurso (varchar)	Código utilizando para identificar o curso	EI24101
AnoLectivo (varchar)	Ano em que o aluno inscreve no curso	2008/2009

Tabela de Facto da Inscrição na Disciplina

A Tabela 10 contém dois tipos de atributos, as métricas ou factos do modelo multidimensional e as chaves estrangeiras que a ligam a cada uma das tabelas de dimensão. Este processo de negócio produz as informações úteis acerca da inscrição na disciplina de um aluno na instituição.

Tabela 10 Descrição de Facto Inscrição na Disciplina

DataInscricao (date)	Chave primária referente à dimensão de InscriçãoDisciplina que permite registar a data de inscrição na disciplina
NumAluno (varchar)	Chave estrangeira referente à dimensão do Aluno. Indica-se aos alunos que realizaram as inscrições
CodCurso (varchar)	Chave estrangeira referente à dimensão do Curso. Indica-se as informações sobre os cursos ministrados na instituição
CodDisciplina (varchar)	Chave estrangeira referente à dimensão Disciplina que se indica as características das disciplinas na instituição
NumProfessor (varchar)	Chave estrangeira referente à dimensão do Professor que tem duas funções importantes como ser professor responsável e ser professor que leciona a unidade curricular. Permite também para registar a informação pessoal do professor

AnoLectivo (varchar)	É uma medida em que permite identificar o ano lectivo em que os alunos realizaram as inscrições na disciplina
----------------------	---

Através das dimensões criadas anteriormente, vamos construir o modelo dimensional para o processo de negócio de Inscrição na Disciplina, como apresentado na Figura 13.

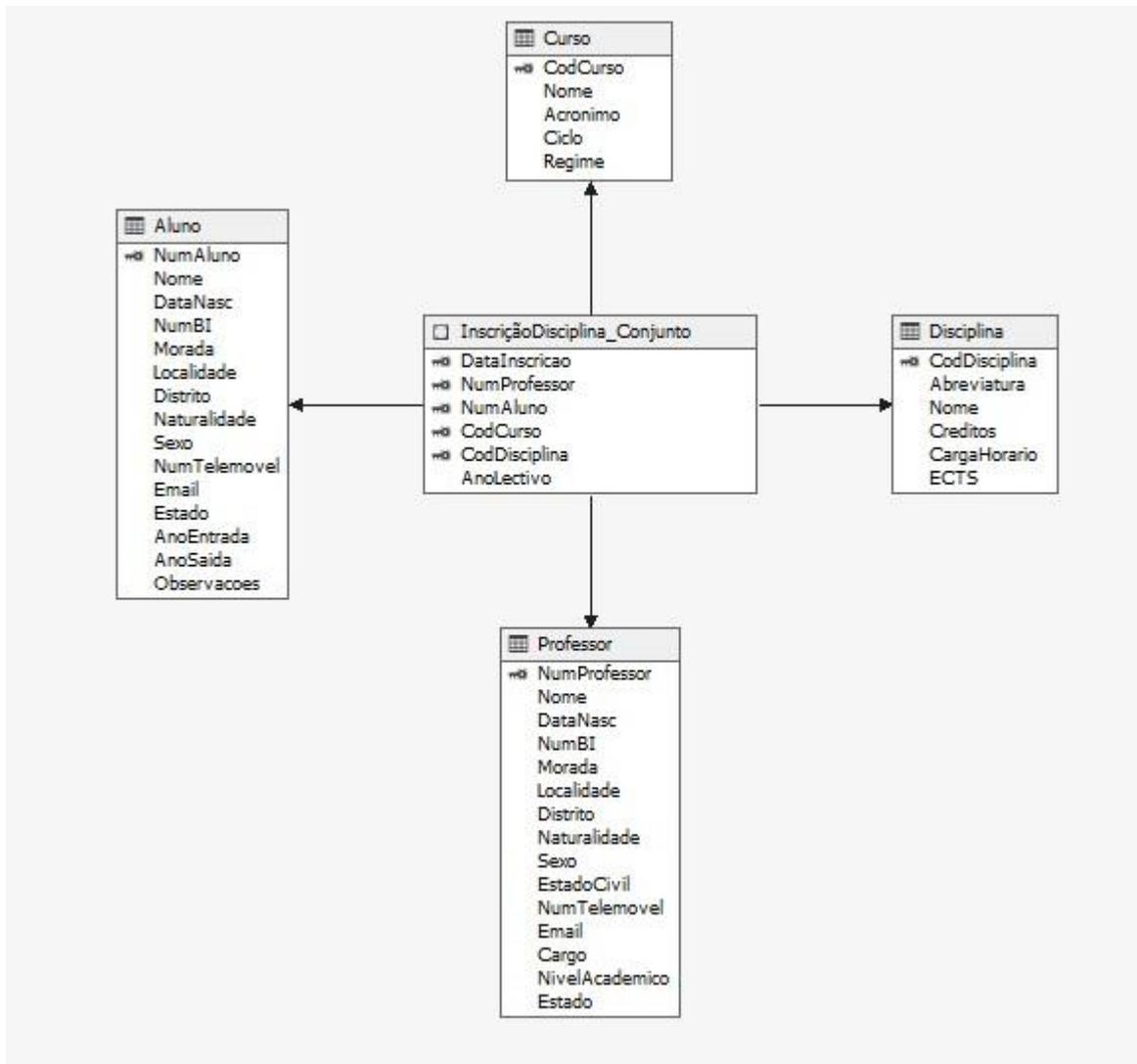


Figura 13 Modelo Dimensional para o Processo de Negócio Inscrição na Disciplina

Tabela de Facto da Pauta

Esta Tabela 11 consiste por dois tipos de atributos que são as métricas ou factos do modelo multidimensional e as chaves estrangeiras que a ligam a cada uma das tabelas de dimensão. Um dos processos de negócios que permite avaliar os alunos é através das notas.

Tabela 11 Descrição de Facto Pauta

Data (date)	É a chave primária desta tabela de dimensão, onde indica a data de produção da nota
-------------	---

AnoLectivo (varchar)	Chave estrangeira referente à dimensão Reingresso
Ano (varchar)	Chave estrangeira referente à dimensão AnoSemestre onde permite identificar o ano de estudo que aluno está, por exemplo: 1º Ano ou 2º Ano
CodDisciplina (varchar)	Chave estrangeira referente à dimensão Disciplina
NumAluno (varchar)	Chave estrangeira referente à dimensão Aluno
CodCurso (varchar)	Chave estrangeira referente à dimensão Curso
Nota (varchar)	É uma medida que se indica à avaliação da nota do aluno

De seguida, apresenta também o modelo dimensional para o processo de negócio da Avaliação_Pautas como apresentado na Figura 14.

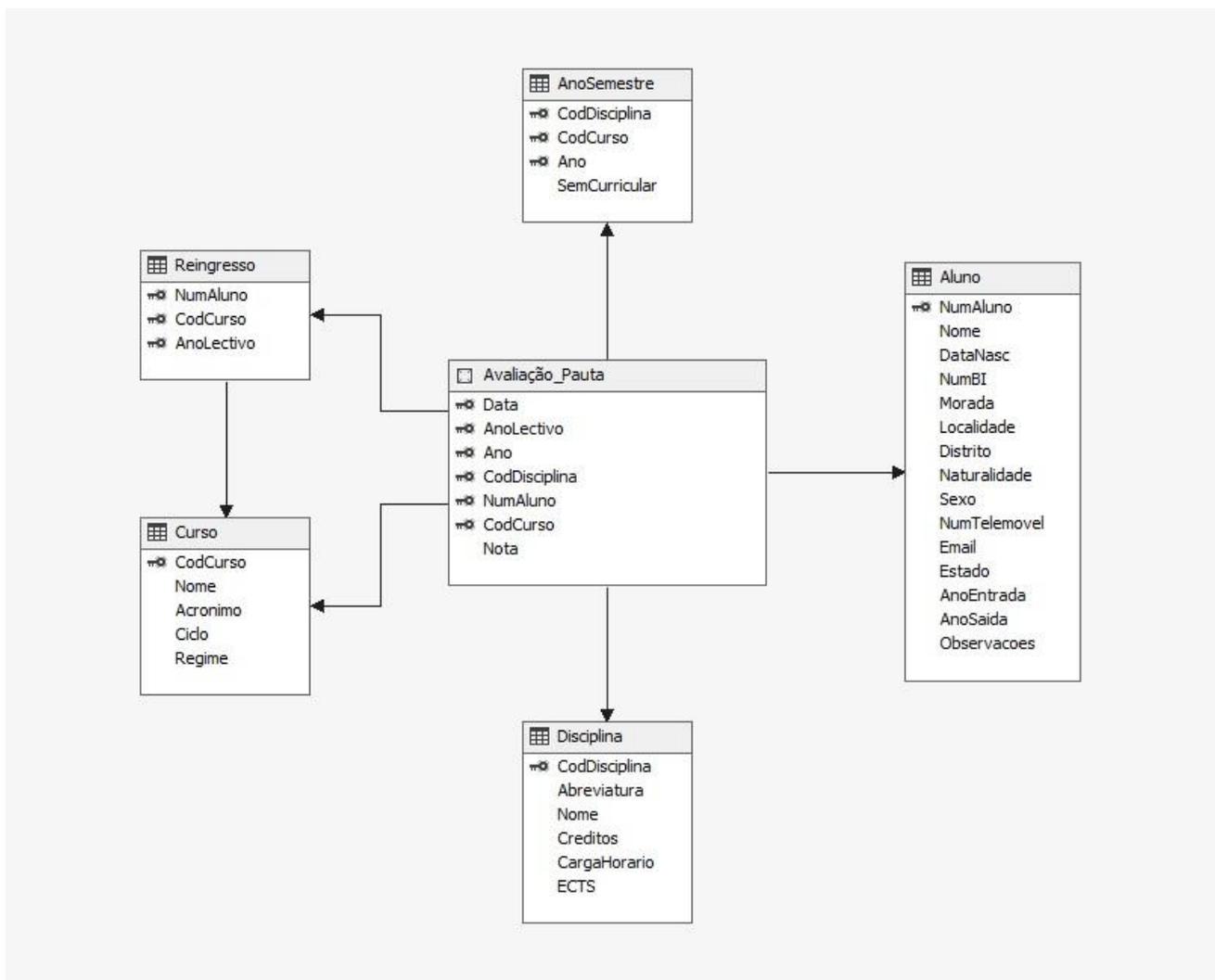


Figura 14 Modelo Dimensional para o Processo de Negócio Avaliação_Pauta do Aluno

4.6 Análise Multidimensional e Relatórios

No desenvolvimento de DW, o cubo de dados representa um conjunto de detalhes de informações. Um cubo OLAP é uma estrutura de dados que supera limitações de bases de dados relacionais, proporcionada uma rápida análise de dados. Sabendo que, esse cubo não é mais do que uma estrutura composta por tabelas de dimensões e por uma tabela de factos. As dimensões integram os eixos do cubo correspondendo aos atributos descritores do âmbito analisado sendo associados os respetivos dados ou medidas da tabela de factos. Os cubos OLAP podem exibir e somar grandes volumes de dados, embora também forneçam aos utilizadores acesso pesquisável a quaisquer pontos de dados. Os dados possam ser acumulados e analisados conforme a necessidade para tratar da maior variedade de questões relevantes à área de interesse do utilizador.

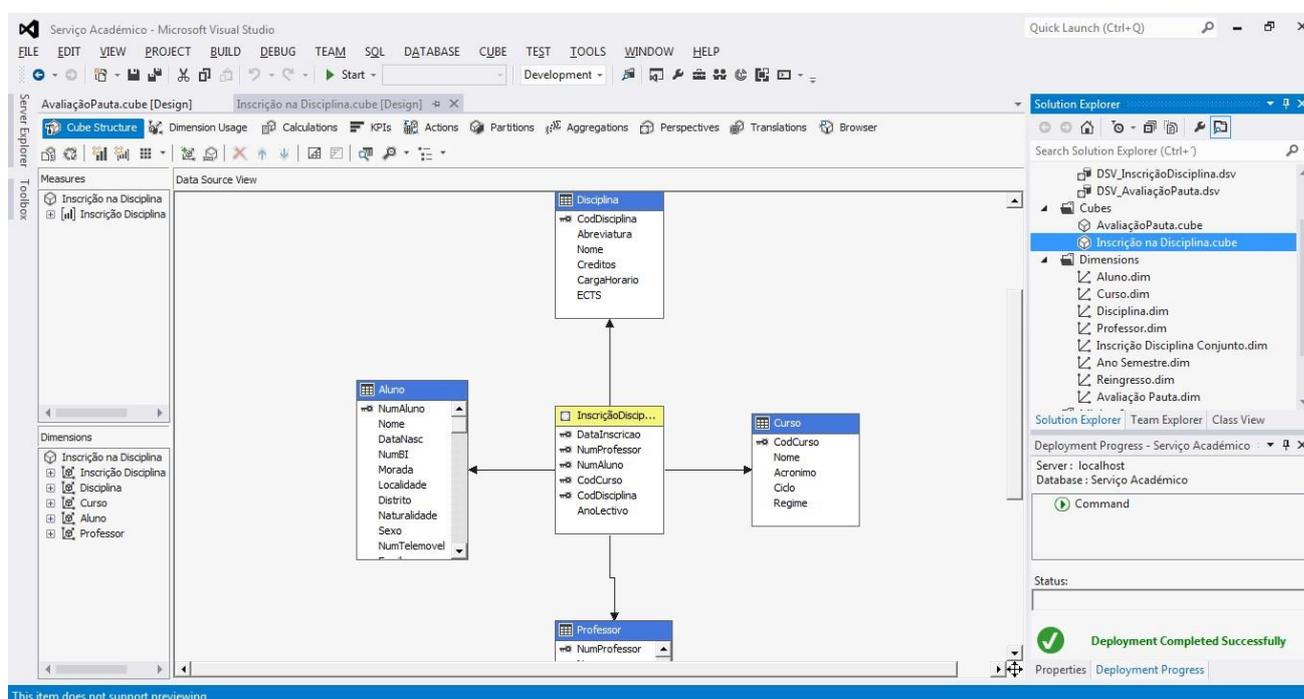


Figura 15 Cubo de Dados Gerado no BIDS

O acesso à informação é efetuado através da ferramenta do SQL Server Analysis Services, permite criar cubos que corresponde as necessidades de consultas, veja na Figura 15 e Figura 16. A elaboração de cubo é começado por ter primeiro à informação de fonte de dados dimensional que já foi executada antes. Logo em seguida, passou-se à seleção das tabelas de dimensões e tabela de facto que são criados e executáveis para serem introduzidas no cubo. Após de criados, implementados e executados, os cubos ficam disponíveis para serem acedidos para análise das suas informações. Desta vez, organizar bem as informações necessárias de maneira como quiser através dos cubos gerados, para que serem visualizados no relatório.

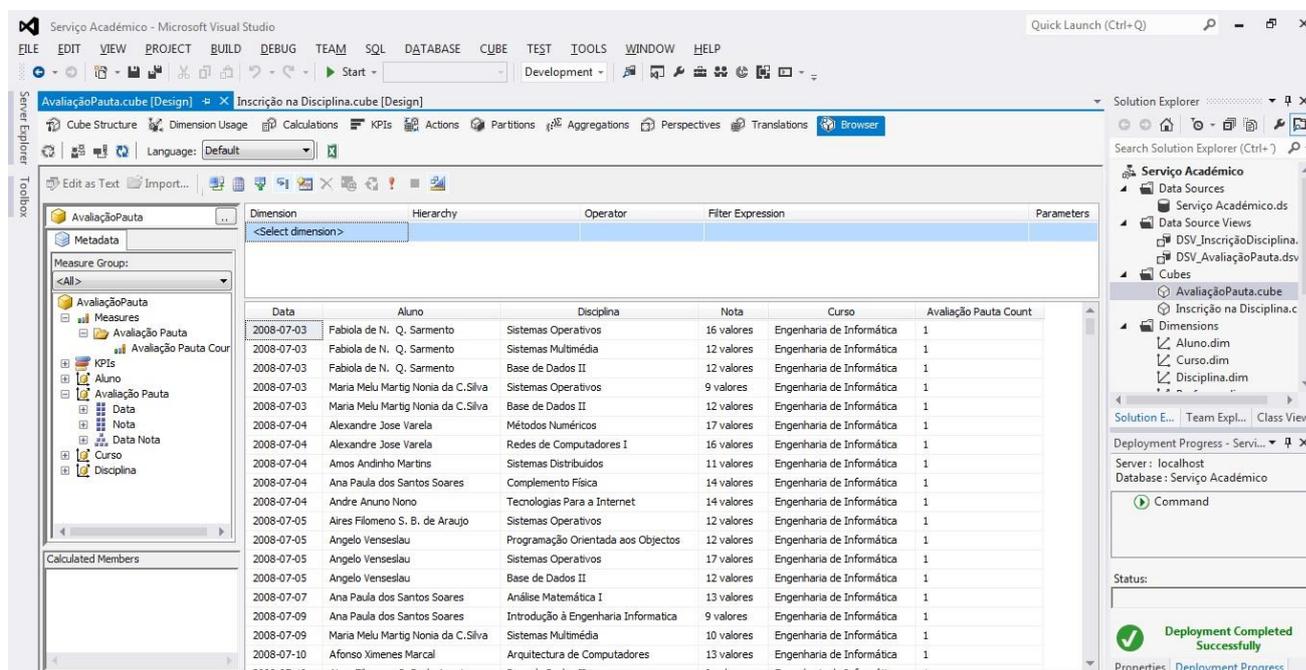


Figura 16 Browse do Cubo de Dados para relatório

Há varias vantagens que um DW traz à atividade de uma organização, o que mais valia de constatar é a facilidade de extração de relatórios integrados bem como a diminuição de tempo despendido na sua construção. Sabia que, o DW é efetuado por diversas aplicações operacionais que tem outros tipos de suporte, e que seja unificar, é necessário seguir por alguns processos, nomeadamente, o processo de ETL. Mesmo que passando por alguns tempos a demorar na construção, por fim encontra-se uma informação esperada.

Uma solução implementada neste trabalho de DW para a disponibilização de relatórios foi efetuada através do Report Builder 3.0 à plataforma de BI, como indicado na Figura 17, Figura 18, Figura 19 e Figura 20. É uma ferramenta para o acompanhamento de resultados e tomadas de decisões quando ligada através do servidor ao Analysis Service, onde as informações são atualizadas automaticamente baseando-se no processamento de cubos OLAP existentes.

Curso	Disciplina	Professor	Aluno	2008/2009	2009/2010	2011/2012	Total		
Engenharia de Informática	Total			75	199	430	704		
	Álgebra Linear e Geometria Analítica	Total		1	1	23	25		
		Zulmira Ximenes da Costa, M.Est	Total		1	1	23	25	
			Adriano Tomas da Costa Cabral					1	1
			Anani Wele Soares					1	1
			Anita da Costa Lemos Boavida					1	1
			Benezato da Costa					1	1
			Carlos Amandio Belo		1				1
			Cipriana Julia da C. J. S.					1	1
			Domingos Soares					1	1
			Elizita Pinto					1	1
Ersilia Vieira Cabral						1	1		
Feliciano dos Santos					1	1			
Fernao Antonio Lopes M. Mouzinho					1	1			
Gabriela Nahak Fernandes					1	1			

Figura 17 Inscrição do Aluno na Disciplina por Ano Letivo

Com a evolução da tecnologia, tornou-se a informação mais-valia, bem organizada e centralizada num local através dos recursos de tecnologia adequados. As informações serão distribuídas com diversas tecnologias que interligam em rede para que ter acesso em qualquer lugar. Dada, este trabalho não seja implementado um portal para disponibilizar as informações por via *online* aos utilizadores, mas sim, por apenas de relatórios simples.

Os relatórios permitem disponibilizar as informações de uma forma bem organizada para que os responsáveis possam analisar e tomar decisão correta. No que diz respeito às análises que serão realizadas neste estudo de caso, existem duas áreas que claramente importante, são: a inscrição do aluno na disciplina e pautas de avaliação do aluno. O primeiro, como indica na Figura 17, a criação do relatório inscrição do aluno na disciplina, é feita por processo dimensional OLAP para analisar os resultados. A sua análise é efetuada por calcular a quantidade total de alunos que inscrevem em cada disciplina e varia por ano letivo. As informações obtidas através das tabelas de dimensões compostas por Curso, Disciplina, Professor, Aluno e uma tabela de Facto de Inscrição na Disciplina. Uma vez que, a informação gerada é parametrizada por estado do aluno “Activo e Finalista”, e não se trata do aluno inativo, pelo que o aluno foi expulso. Assim, podemos analisar de forma real, as quantidades dos alunos que inscreveram numa disciplina por período de tempo.

No quadro apresentado na Figura 17, conseguimos observar, por exemplo, a quantidade total dos alunos ativos e finalistas que inscreveram na disciplina ou seja no curso no período de 2008/2009 são

75 alunos, de 2009/2010 são 199 alunos, de 2011/2012 são 430 e total são 704 alunos. É uma análise bem simples e direcionadas para a visão global da instituição ao longo da sua existência.

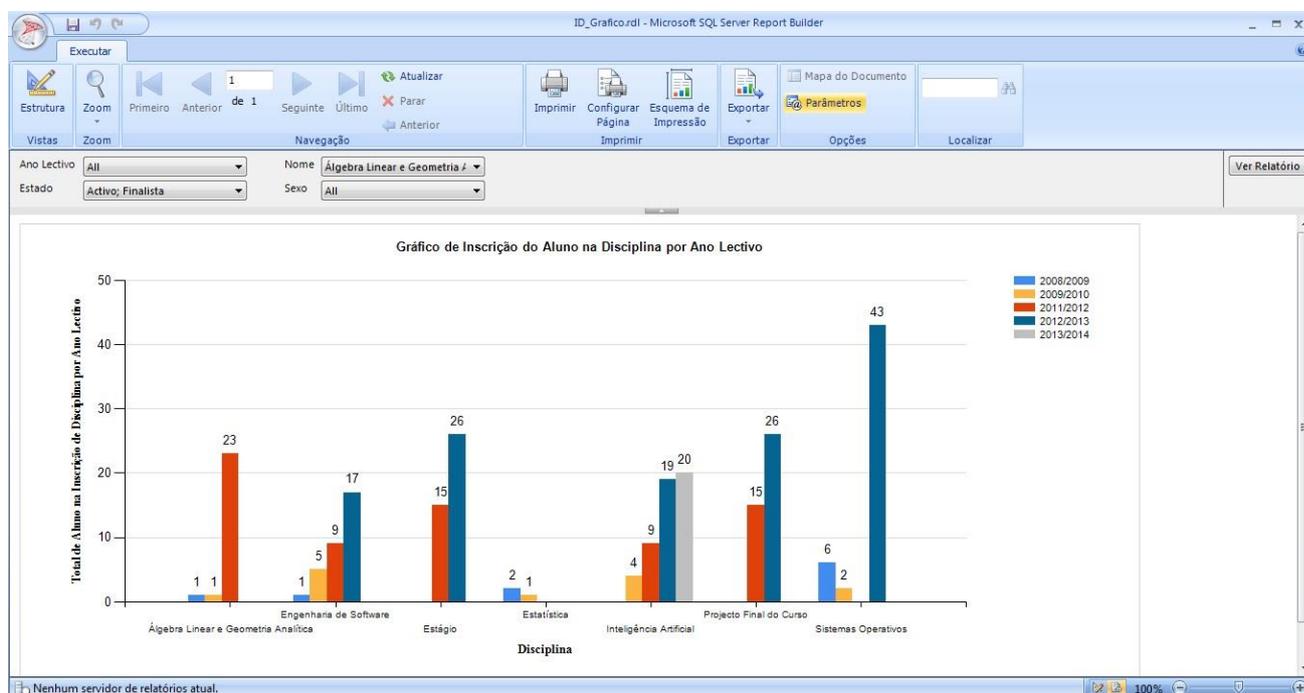


Figura 18 Gráfico de Inscrição do Aluno na Disciplina por Ano Letivo

Demostrou-se na Figura 18, uma gráfica simples que podendo analisar o número total do aluno que foi inscrito em cada disciplina e variado por ano letivo. Desta vez, é considerado 43 alunos que inscreveram na disciplina de Sistemas Operativos no ano letivo de 2012/2013.

Segundo, é apresentada na Figura 19 e Figura 20, o relatório sobre pautas de avaliação do aluno. Esse relatório permite disponibilizar as informações necessárias baseadas nas tabelas dimensionais, composto por Curso, Disciplina, Aluno e uma tabela de facto Pauta. A análise é efetuada por calcular a quantidade do aluno por pauta na disciplina. As notas disponíveis são de 8 até 18 valores em que correspondem aos alunos ativo e finalistas para todo período de tempo.

Curso	Disciplina	Aluno	10 valores	11 valores	12 valores	13 valores	14 valores	15 valores	16 valores	17 valores	18 valores	19 valores	Total
Engenharia de Informática	Total		118	15	377	48	94	81	90	122	43	17	175
	Álgebra Linear e Geometria Analítica	Total	5		5	1	1	2	2	2	1	1	5
		Adriano Tomas da Costa Cabral											1
		Anani Wele Soares	1										1
		Anita da Costa Lemos Boavida	1										1
		Benezato da Costa			1								1
		Carlos Amandio Belo			1								1
		Cipriana Julia da C. J. S.						1					1
		Domingos Soares	1										1
		Elizita Pinto								1			1
		Ersilia Vieira Cabral										1	1
		Feliciano dos Santos							1				1
		Fernao Antonio Lopes M. Mouzinho										1	1
		Gabriela Nahak Fernandes			1								1
		Guilhermino da Costa Soares										1	1
		Idelfonso Ximenes	1										1
		Jose J. F do Rego Miranda						1					1
		Juvenica D.J. Pereira										1	1
		Ludigerio Jorge Soares	1										1
		Manuel da Silva Marques				1							1
		Martha Lopes									1		1
		Nelson David Fernandes							1				1

Figura 19 Pautas de Avaliação do Aluno

A gráfica demonstrada na Figura 20, indicou-se a maioria dos alunos que tiveram uma nota de 12 valores na disciplina de Sistemas Multimédia, Inteligência artificial, Estágio e entre outras. E também podemos observar e analisar que ao longo da existência do curso, a minoria dos alunos que obterem boas notas. De todo que foi apresentado, esta é uma das análises bem detalhada em qualquer nível para suporte a tomada de decisão académica.

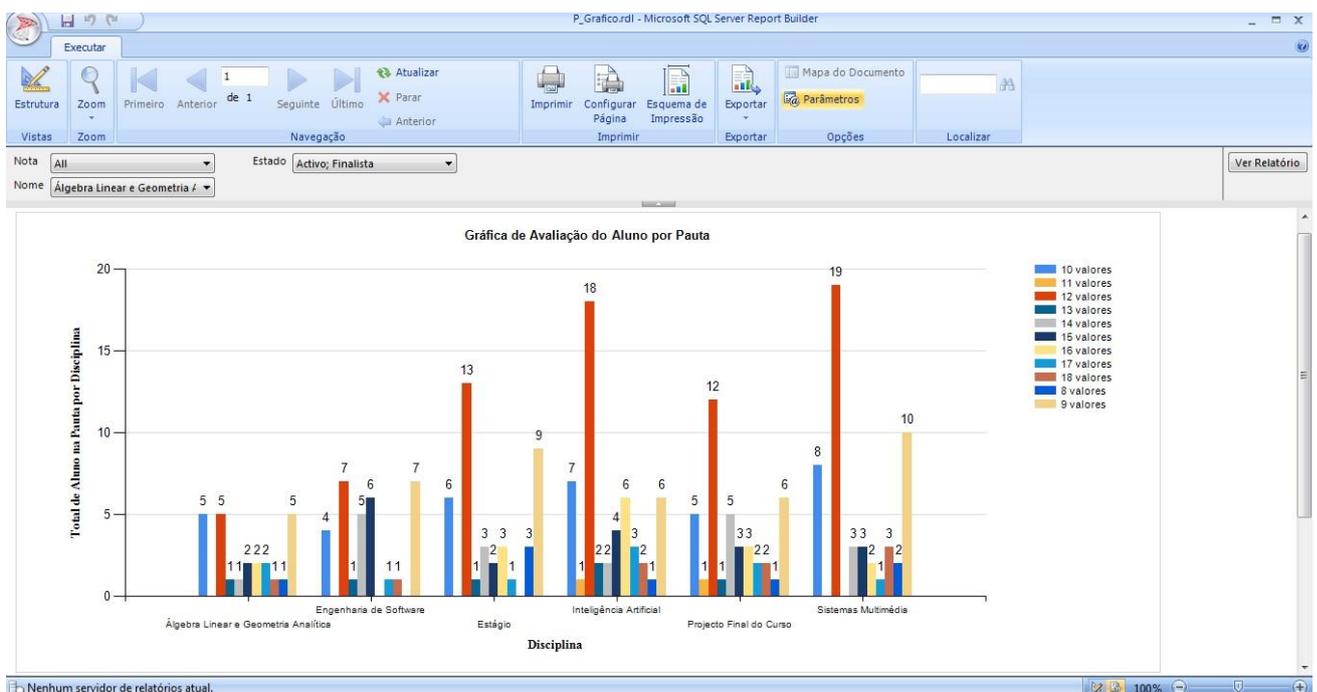


Figura 20 Gráfica de Avaliação do Aluno por Pauta

Capítulo 5 Conclusões e Trabalhos Futuros

Conclusões

A efetuar uma análise e tomar decisões, fazendo uso de grandes quantidades de informação é inviável para o ser humano, pelo que o recurso à tecnologia com ferramentas computacionais apropriadas é fundamental. Com a evolução da tecnologia torna-se imprescindível o desenvolvimento de sistemas que possam interpretar as informações necessárias em cada passo da aplicação. Neste contexto, a tecnologia de DW mostra-se interessante para as instituições que geram e possuem grandes volumes de dados acumulados ao longo da sua existência, para que lhes seja possível tomar decisões de forma rápida e eficaz.

O desenvolvimento da tecnologia de sistemas de DW no ensino superior de Timor é muito recente, se comparado com as instituições do ensino superior em países desenvolvidos. O principal problema encontrado, foi, não terem sido utilizados sistemas de informação baseados em tecnologias de informação e comunicação para apoiar a gestão da instituição.

Este estudo contribui com trabalho específico na área de DW, para melhorar o sistema de informação da UNTL. Somente uma pequena parte da informação relativa à UNTL está em suporte digital sendo que desta maioria a aplicação usada para armazenar informação foi o Excel em vez de outras aplicações mais apropriadas. Tratou-se portanto duma questão importante implementar uma plataforma dos serviços académicos que permitisse a construir do DW da instituição com informação correta e apropriada.

Dadas as limitações de recursos humanos e tecnologias e deficiente acesso às informações existentes, é de esperar este que trabalho de dissertação seja útil na concretização futura de um grande projeto da UNTL.

Bibliografia

- Albino, N. L. (2010). *Estudo de Implementação de um Data Warehouse para uma Instituição de Ensino Superior*. Universidade de Évora: Dissertação.
- Alex dos Santos Vieira, R. P. (s.d.). *Data Warehouses-Uma Introdução*. Obtido em 25 de Fevereiro de 2012, de http://www.fij.br/revista/arq/vol01_01/datawarehouse.pdf
- BURTESCU, E. (n.d.). *DATA WAREHOUSE AND DATA MINING – NECESSITY OR USELESS*. Retrieved 2012 йил 19-Novembro from <http://economic.upit.ro/repec/pdf/l1.pdf>
- Caldeira, C. P. (2008). *Data Warehouse: Conceitos e Modelos* (Edição Sílabo ed.). Lisboa.
- Camejo, C. M. (2010). *Desenvolvimento de um Sistema de Data Warehouse para Consolidação da Informação num Grupo Empresarial*. Tese de Mestrado, Universidade de Évora, Évora.
- Cleomir Kuhnen, V. A. (s.d.). Obtido em 20 de Janeiro de 2013, de Vantagens Obtidas Com a Utilização da Implementação do Data Warehouse: Um Estudo de Caso na Unimed de Francisco Beltrão: <http://www.ead.fea.usp.br/semead/8semead/resultado/trabalhosPDF/20.pdf>
- Daniele Del Bianco Hokama, D. C. (2004). *A Modelagem de Dados no Ambiente Data Warehouse*. Retrieved 2012 йил 22-Setembro from <http://meusite.mackenzie.com.br/rogerio/tgi/2004ModelagemDW.pdf>
- Di Domenico, J. A. (2001). *Definição de um Ambiente Data Warehouse em uma Instituição de Ensino Superior*. Obtido em 11 de Setembro de 2014, de <http://repositorio.ufsc.br/xmlui/handle/123456789/79520>
- Donizete Carlos Bruzarosco, A. V. (2000). *Criando data warehouse com o modelo dimensional*. Retrieved 2012 йил 22-Setembro from <http://periodicos.uem.br/ojs/index.php/ActaSciTechnol/article/download/3099/2225>
- Fujiwara, D. K. (2006). *Data Warehouse Como Instrumento de Suporte à Avaliação Académica*. Tese de Mestrado, Universidade de Brasília (UnB), Brazil.
- Goede, R. (2005). *CHAPTER 4 DATA WAREHOUSING*. Retrieved 2012 йил 25-Fevereiro from <http://upetd.up.ac.za/thesis/available/etd-05132005-080727/unrestricted/04chapter4.pdf>
- Inmon, W. H. (2002). *Building the Data Warehouse* (Third Edition ed.). (J. W. Sons, Ed.) New York: Wiley Publishing.
- Inmon, W. H. (2005). *Building the Data Warehouse(Fourth Edition)*. New York: Wiley Publishing.
- Jiawei Han, M. K. (2001). *Data Mining - Concepts and Techniques* (Second Edition ed.). New York, USA: Morgan Kaufmann Publishers.
- Júnior, H. L. (n.d.). *Data Warehouse Aplicado em Instituições de Ensino Superior*. Retrieved 2012 йил 24-Outubro from DevMedia: <http://www.devmedia.com.br/DATA-WAREHOUSE-APLICADO-EM-INSTITUIC%C3%95ES-DE-ENSINO-SUPERIOR.html/5538>
- Kimball, R. (2004). *The Data Warehouse ETL Toolkit*. Canadá: Wiley Publishing.
- Kimball, R. (2008). *The Data Warehouse Lifecycle Toolkit - Practical Techniques for Building Data Warehouse and Business Intelligence Systems* (Second Edition ed.). Canadá: Wiley Publishing.
- Magalhães, A. (2013). *SQL Server 2012 - Curso Completo*. Lisboa: FCA.
- Microsoft. (2012). *SQL Server 2012 Tutorials: Analysis Services - Multidimensional Modeling*. Microsoft Corporation.
- Moreira, E. (27 de 01 de 2006). *Modelo Dimensional para Data Warehouse*. Obtido em 16 de Agosto de 2012, de iMasters: <http://imasters.com.br/artigo/3836/gerencia-de-ti/modelo-dimencional-para-data-warehouse>
- Oliveira, J. L. (2009). *Sistema de Business Intelligence "Sistema de Apóio a Gestão para Tomadas de Decisões Inteligentes"*. Retrieved 2012 йил 19-Novembro from Slideshare: <http://www.slideshare.net/joaoluiciooliveira/monografia-business-intelligence>
- Ponniah, P. (2002). *Data Warehousing Fundamentals: A Comprehensive Guide for IT Professionals*. New York: John Wiley & Sons.

- Ponniah, P. (2010). *Data Warehousing Fundamentals for IT Professional(Second Edition)* (Second Edition ed.). (J. W. Sons, Ed.) Canadá: John Wiley & Sons.
- Ralph Kimball, M. R. (2002). *The Data Warehouse Toolkit(Second Edition) The Complete Guide to Dimensional Modeling*. Canadá: Wiley Computer Publishing.
- Raph Kimball, L. R. (1998). *The Data Warehouse Lifecycle Toolkit - Expert Methods for Designing, Developing and Deploying Data Warehouses*. Canadá: John Wiley & Sons.
- Somantri, O. (s.d.). Obtido em 25 de Novembro de 2012, de Aplikasi Data Warehouse Pada SQL Server: http://www.4shared.com/document/f7vbag4z/Aplikasi_Data_Warehouse_dengan.html?aff=7637829
- Sousa, R. S. (1 de Fevereiro de 2010). *Wordpress*. Obtido em 20 de Agosto de 2014, de Guias e Dicas: <http://ruisalgueirosousa.wordpress.com/2010/02/01/bus-matrix-uma-breve-explicacao>
- Subhan. (s.d.). *Mengenal Dasar-Dasar SQL Server 2008 R2 Analysis Service*. Obtido em 25 de Novembro de 2012, de <http://blog.stikom.edu/rachmawati/files/2012/09/Mengenal-Dasar-Dasar-SQL-Server-2008R2-Analysis-Service.pdf>
- Suryanto, W. D. (2010). *Pengembangan Data Warehouse dan Aplikasi OLAP Data Tracer Study Alumni IPB Berbasis Web Menggunakan Microsoft Business Intelligence*. Obtido em 11 de Setembro de 2014, de http://wdwisuryanto.staff.ipb.ac.id/files/2012/01/Makalah_Praseminar1.pdf
- William H. Inmon, C. I. (1999). *Exploration Warehousing*. New York: John Wiley & Sons.

Anexos

1. Programa em java, criar uma nova folha de cálculo (Excel) com função de fazer separação entre o “Local” e a “Data” na mesma coluna dentro dos dados de origens.

```
package Ler_Ficheiro_Excel;
import com.sun.rowset.internal.Row;
import java.io.File;
import java.io.FileOutputStream;
import java.io.IOException;
import java.util.StringTokenizer;
import java.util.logging.Level;
import java.util.logging.Logger;
import jxl.Cell;
import jxl.CellType;
import jxl.Sheet;
import jxl.format.Border;
import jxl.Workbook;
import jxl.read.biff.BiffException;
import java.util.Date;
import jxl.*;
import jxl.format.BorderLineStyle;
import jxl.write.*;
import org.apache.poi.hssf.usermodel.*;
import org.apache.poi.hssf.util.HSSFColor;

public class Excel_Final {
    public static void main(String[] args) throws IOException, BiffException, WriteException
    {
        Workbook = Workbook.getWorkbook(new File("_DepInformática_Impar.xls"));
        WritableWorkbook copy = Workbook.createWorkbook(new File("output.xls"),
workbook);
        WritableCellFormat cellFormat = new WritableCellFormat();
        cellFormat.setBorder(Border.ALL, BorderLineStyle.THIN);

        int numeroSheet = 10;
        for (int i=0; i<numeroSheet; i++){
            WritableSheet sheet2 =copy.getSheet(i);
            sheet2.insertColumn(5);

            Sheet = workbook.getSheet(i);
            /* Ciclo que calcula o tamanho da tabela. A ideia é percorrer toda a tabela do
            número (ultima), pois esta coluna vai estar sempre preenchida.
            */
            int contadorTab = 0;
```

```

for (int contadorTabela=6; contadorTabela >0; contadorTabela++){
    Cell a = sheet2.getCell(0, contadorTabela);
    String stringa = a.getContents();
    if (stringa==""){
        contadorTab=contadorTabela;
        System.out.println(contadorTabela);
        break;
    }
}
}

```

/* Vai percorrer toda a primeira coluna para identificar o conteúdo: se tiver virgula, é porque há uma localidade e uma data. se não tiver virgula, vai ver qual o primeiro caracter: se for número é porque só uma data e muda-se a data de coluna. Se não tiver nada ou só a localidade, não se mexe. Não está a funcionar para os casos que não tiver localidade nem data.

A ideia também será com o tamanho da tabela calculado anteriormente, saber até que célula vamos analisar e modificar.

```

*/
for (int contador=6;contador <contadorTab ;contador++){
    Cell a1 = sheet.getCell(4,contador);
    String stringa1 = a1.getContents();
    System.out.println(stringa1);
    String s[] = null;
    // Verifica se tem vírgula, se sim, separa o conteúdo e mete na célula correta.
    if(stringa1.contains(", ")){
        s = stringa1.split(", ");

        Label = new Label(4, contador, s[0],cellFormat);
        sheet2.addCell(label);

        Label label2 = new Label(5, contador, s[1], cellFormat);
        sheet2.addCell(label2);
        System.out.println(s[0]);
        System.out.println(s[1]);
    }
    else {
        try {
            /* Se não tem virgula, verifica se o primeiro caracter é numero, se sim é porque é data e vai para o sitio correto.
            */
            char [] primeiraPosicao = stringa1.toCharArray();
            if (primeiraPosicao [0] >= '0' && primeiraPosicao [0] <='9'){

                Label = new Label(4, contador, "",cellFormat);
                sheet2.addCell(label);
                Label label2 = new Label(5, contador, stringa1,cellFormat);
                sheet2.addCell(label2);
            }
        } catch (ArrayIndexOutOfBoundsException e) {
            System.out.println("celula sem localidade nem data");
            Label label2 = new Label(5, contador, "",cellFormat);

```

```

        sheet2.addCell(label2);
    }
}
}
// Renomeia os titulos das colunas
WritableCellFormat cellFormat2 = new WritableCellFormat();
if (i+1<numeroSheet){
    cellFormat2.setBorder(Border.ALL, BorderLineStyle.THIN);
    cellFormat2.setAlignment(jxl.format.Alignment.CENTRE);
    cellFormat2.setBackground(jxl.format.Colour.YELLOW);
}
else {
    cellFormat2.setBorder(Border.ALL, BorderLineStyle.THIN);
    cellFormat2.setBorder(Border.TOP, BorderLineStyle.DOUBLE);
    cellFormat2.setAlignment(jxl.format.Alignment.CENTRE);
    cellFormat2.setBackground(jxl.format.Colour.GREY_25_PERCENT);
}

Label label3 = new Label(4, 5, "LOCAL",cellFormat2);
sheet2.addCell(label3);
Label label4 = new Label(5, 5, "DATA",cellFormat2);
sheet2.addCell(label4);
System.out.println(i+" sheet concluido");
}
copy.write();
copy.close();
}
}
}

```

2. Script de base de dados em SQL Server

```

CREATE TABLE tblProfessor (
    CodProfessor varchar(10) PRIMARY KEY NOT NULL,
    Nome varchar(50) NULL,
    DataNasc date NULL,
    NumBI varchar(10) NULL,
    Morada varchar(50) NULL,
    Localidade varchar(50) NULL,
    Distrito varchar(50) NULL,
    Naturalidade varchar(50) NULL,
    CodPostal varchar(10) NULL,
    Sexo char(1) NULL,
    Estado_civil varchar(50) NULL,
    NumTelemovel varchar(8) NULL,
    Email varchar(50) NULL,
    Cargo varchar(50) NULL,
    Nivel_academico varchar(50) NULL,
    Estado varchar(50) Null
)

```

```

CREATE TABLE tblDisciplina (
    CodDisciplina varchar(10) PRIMARY KEY NOT NULL,
    Abreviatura varchar(10) NULL,
    Nome varchar(50) NULL,
    Peso varchar(10) NULL,
    Creditos varchar(10) NULL,
    CargaHorario varchar(50) NULL,
    ECTS varchar(10) NULL
)

CREATE TABLE tblCurso (
    CodCurso varchar(10) PRIMARY KEY NOT NULL,
    Nome varchar(50) NULL,
    Acronimo varchar(10) NULL,
    Ciclo varchar(50) NULL,
    Regime varchar(50) NULL
)

CREATE TABLE tblAnoSemestre (
    CodDisciplina varchar(10) FOREIGN KEY REFERENCES tblDisciplina(CodDisciplina) NOT
NULL,
    CodCurso varchar(10) FOREIGN KEY REFERENCES tblCurso(CodCurso) NOT NULL,
    Ano varchar(10) PRIMARY KEY NOT NULL,
    SemCurricular varchar(50) NULL
)

CREATE TABLE tblReingresso (
    NumAluno varchar(10) FOREIGN KEY REFERENCES tblAluno(NumAluno) NOT NULL,
    CodCurso varchar(10) FOREIGN KEY REFERENCES tblCurso(CodCurso) NOT NULL,
    Anolectivo varchar(10) PRIMARY KEY NOT NULL
)

CREATE TABLE tblInscricaoDisciplina (
    NumAluno varchar(10) FOREIGN KEY REFERENCES tblAluno(NumAluno) NOT NULL,
    CodDisciplina varchar(10) FOREIGN KEY REFERENCES tblDisciplina(CodDisciplina) NOT
NULL,
    Anolectivo varchar(10) NULL,
    DataInscricao date PRIMARY KEY NOT NULL
)

CREATE TABLE tblPauta (
    NumAluno varchar(10) FOREIGN KEY REFERENCES tblAluno(NumAluno) NOT NULL,
    CodDisciplina varchar(10) FOREIGN KEY REFERENCES tblDisciplina(CodDisciplina) NOT
NULL,
    Data date PRIMARY KEY NOT NULL,
    Nota varchar(10) NULL
)

```

```

CREATE TABLE tblLeciona (
    CodDisciplina varchar(10) FOREIGN KEY REFERENCES tblDisciplina(CodDisciplina) NOT
NULL,
    CodProfessor varchar(10) FOREIGN KEY REFERENCES tblProfessor(CodProfessor) NOT
NULL,
)

```

3. Desenho do modelo Entidade Relacional

