



Universidade de Évora

Escola de Ciências Sociais

Mestrado em Gestão

Área de Especialização - Organização e Sistemas de Informação

Construção de um protótipo de *Data Warehouse* como suporte ao sistema de informação numa instituição de ensino superior

Trabalho de Projecto apresentado para a obtenção do grau de Mestre em Gestão

Autor

Joaquim Pombo Da Silva Dias

Orientador:

Professor Doutor Paulo Resende da Silva

Fevereiro de 2012

Mestrado em Gestão

Área de Especialização - Organização e Sistemas de Informação

Construção de um protótipo de *Data Warehouse* como suporte ao sistema de informação numa instituição de ensino superior

Trabalho de Projecto apresentado para a obtenção do grau de Mestre em Gestão

Autor

Joaquim Pombo Da Silva Dias

Orientador:

Professor Doutor Paulo Resende da Silva

"Um *Data Warehouse* não se pode comprar, tem que se construir".

(Anónimo)

AGRADECIMENTOS

Desejo manifestar os meus agradecimentos e gratidão a todos os que directa ou indirectamente contribuíram para a realização deste trabalho.

Ao meu orientador Professor Doutor Paulo Resende Silva; por me ter aceite como seu orientando, pelas suas opiniões, pela permanente disponibilidade, pelo seu estímulo precioso no processo de elaboração deste trabalho, pelo apoio manifestado ao longo de todo este período do curso que se revelou de valor incalculável para a concretização deste desafio e por sempre ter acreditado neste trabalho desde o início.

A todos os que foram meus professores e que me deixaram a sua desde a escola primária até ao mestrado.

Aos amigos e colegas de trabalho que me incentivaram na realização deste trabalho, em particular às colegas e amigas Sílvia Silva e Elisabete Ferreira pelo apoio dado desde o primeiro ano do mestrado.

Ao anterior presidente do Instituto Politécnico de Tomar, Dr António Pires da Silva, por ter aceiteado a realização deste trabalho no IPT.

À minha família e, em especial, à minha esposa Ana Ferreira e filho André Pombo, aos quais dedico este trabalho, pelo grande apoio, carinho, compreensão e incentivos permanentes, os quais foram fundamentais para a sua realização e conclusão.

Um obrigado muito especial aos meus pais e avós, por todos os valores transmitidos e que apesar de ausentes estiveram sempre presentes.

RESUMO

Construção de um protótipo de *Data Warehouse* como Suporte ao Sistema de Informação numa Instituição de Ensino Superior

Uma das dificuldades que se verifica na extracção de informação numa organização é a falta de integração dos dados existentes dispersos em diversos formatos: ficheiros de processadores de texto, folhas de cálculo, bases de dados, entre outras fontes. A partir deste problema, este trabalho propõe a estruturação de um modelo de *Data Warehouse* com o objectivo de organizar, armazenar e integrar as informações provenientes de outros formatos e sistemas, numa única base de dados para uma futura utilização no suporte à tomada de decisão. Existem, neste momento, na comunidade de *Data Warehousing* duas principais abordagens, uma preconizada por William H. Inmon, mais centrada nos dados, e outra por Ralph Kimball, mais centrada no projecto.

Assim, com a metodologia proposta foi desenvolvido um caso de estudo com a finalidade de verificar e avaliar a aplicabilidade da metodologia no Instituto Politécnico de Tomar.

Palavras-chave: Sistemas de *Data Warehousing*, *Data Mart*, *Data Warehouse*, Instituições de Ensino Superior, Gestão da Informação.

ABSTRACT

Building of a prototype *Data Warehouse* to Support Information System in a Higher Education Institution

One difficulty that exists in the extraction of information in organizations is the lack of integration of existing data scattered in various formats: word processing files, spreadsheets, databases, among other sources. From this problem, this paper proposes to structure a model of *Data Warehouse* in order to organize, store and integrate information from other systems and formats in a single database for future use in supporting decision making. There are at present in the community of Data Warehousing two main approaches, one advocated by William H. Inmon, more data-centric, and one by Ralph Kimball, more focused on the project.

So with the proposed methodology was developed a case study in order to verify and evaluate the applicability of the methodology at the Polytechnic Institute of Tomar.

Key-words: *Data Warehousing System, Data Mart, Data Warehouse, Higher Education Institutions, Information Management.*

ÍNDICE

LISTA DE TABELAS	VII
LISTA DE FIGURAS	VIII
LISTA DE ABREVIATURAS	IX
GLOSSÁRIO	X
1. INTRODUÇÃO.....	1
1.1. PROBLEMA DE ESTUDO	2
1.2. OBJECTIVOS	3
1.3. METODOLOGIA.....	4
1.4. ESTRUTURA DO TRABALHO	5
2. ENQUADRAMENTO TEÓRICO	7
2.1. <i>DATA WAREHOUSE</i>	7
2.1.1 <i>Conceitos Básicos</i>	7
2.1.2 <i>Características</i>	10
2.1.3 <i>Arquitectura de Data Warehouse</i>	14
2.1.4 <i>Tipos de Implementação</i>	21
2.1.5 <i>Granularidade</i>	24
2.2. METADADOS.....	27
2.3. MODELO DE DADOS	28
2.3.1. <i>Modelo de Dados Entidade Relacionamento (er)</i>	28
2.3.2. <i>Modelo de Dados Dimensional</i>	29
2.3.3. <i>Factos</i>	30
2.3.4 <i>Dimensões</i>	32
2.3.5. <i>Métricas ou medidas</i>	33
2.4. ESQUEMAS DO MODELOS DE DADOS DIMENSIONAL.....	33
2.4.1 <i>Esquema em estrela</i>	34
2.4.2 <i>Esquema em flocos de neve</i>	35
2.4.3 ESQUEMA EM CONSTELAÇÃO.....	35
2.5. <i>DATA MART</i>	36
3. METODOLOGIA DO PROJECTO DW	39
3.1 KIMBALL	39
3.2 INMON	43
3.3 INMON VERSUS KIMBALL.....	47
3.4 <i>MICROSOFT</i>	49
4. ORGANIZAÇÃO DO ENSINO SUPERIOR E O IPT	53
4.1 INTRODUÇÃO	53
4.2 ORGANIZAÇÃO DO ENSINO SUPERIOR	53
4.3 INSTITUTO POLITÉCNICO DE TOMAR	57
4.3.1 <i>Organização interna</i>	58
4.3.2 <i>Oferta formativa</i>	60
4.4 ESTRUTURA ORGANIZACIONAL DO IPT	61
4.4.1 <i>Valores e Missão</i>	61

4.4.2 Objectivos.....	62
4.4.3. Governo e gestão.....	63
4.4.4 Processo estratégico	64
4.4.5 Práticas de Avaliação de Qualidade.....	66
4.5 AVALIAÇÃO EUA	67
4.6 AVALIAÇÃO A3Es	68
5 DESENHO METODOLÓGICO.....	71
5.1 SISTEMAS FONTE EXISTENTES	73
5.2 SISTEMA DE ARMAZENAMENTO	77
5.3 FERRAMENTAS DE EXPLORAÇÃO, MODELAÇÃO E EXPLORAÇÃO.....	78
5.4 ACTIVIDADES PARA A IMPLEMENTAÇÃO DO SISTEMA DE <i>DATA WAREHOUSE</i>	79
5.4.1 Definição de Requisitos	79
5.4.2 Análise e Modelação	82
5.4.2 Desenvolvimento testes e Documentação	97
6. CONCLUSÕES E RECOMENDAÇÕES.....	99
6.1 CONCLUSÕES	99
6.2 TRABALHOS FUTUROS	102
REFERÊNCIAS BIBLIOGRÁFICAS	105
ANEXOS	107
ANEXO A: GUIÃO DE AUTO AVALIAÇÃO DOS CICLOS DE ESTUDOS A3Es.....	109
ANEXO B: CARTA A SOLICITAR AUTORIZAÇÃO PARA A RECOLHA DE DADOS	111
ANEXO C: ALGUNS COMANDOS <i>SQL</i> UTILIZADOS NA EXTRACÇÃO DE DADOS	112
ANEXO D: EXEMPLOS DA INTERFACE DO PROGRAMA	115
ANEXO E: EXEMPLOS DE MODELAÇÃO UTILIZANDO O VISIO	121

LISTA DE TABELAS

TABELA 1 - DEFINIÇÕES DE DATA WAREHOUSE.....	8
TABELA 2 – CARACTERÍSTICAS DO DATA WAREHOUSE	9
TABELA 3-VANTAGENS/DESVANTAGENS DA ARQUITECTURA GLOBAL.....	15
TABELA 4 VANTAGENS/DESVANTAGENS DA ARQUITECTURA <i>DATA MART</i>	17
TABELA 5 - METODOLOGIA PARA DESENVOLVIMENTO DE UM DW (INMON).....	46
TABELA 6 - VANTAGENS METODOLOGIAS INMON E KIMBALL.....	48
TABELA 7 - DESVANTAGENS METODOLOGIAS INMON E KIMBALL.....	48
TABELA 8- DISTRIBUIÇÃO COMUNIDADE ACADÉMICA IPT.....	61
TABELA 9- TABELA MATRICULAS POR CICLO/ANO LECTIVO	82
TABELA 10 - CANDIDATURAS, MATRICULAS, VAGAS E COLOCAÇÕES	84
TABELA 11 - EVOLUÇÃO DO NÚMERO DE CANDIDATURAS	85
TABELA 12 - NÚMERO DE ESTUDANTES DO 1º CICLO GÉNERO.....	86
TABELA 13 - EVOLUÇÃO DO NÚMERO DE ESTUDANTES	86
TABELA 14 - NUMERO DE ESTUDANTES POR FAIXA ETÁRIA	88
TABELA 15 - NÚMERO DE ESTUDANTES POR NACIONALIDADE	90
TABELA 16 - NUMERO DE ESTUDANTES POR REGIÕES	91
TABELA 17 - NUMERO DE ESTUDANTES POR ESCOLA/ANOS CURRICULARES NUM ANO LECTIVO	91
TABELA 18 - ESTUDANTES POR TIPO	92
TABELA 19 - NÚMERO DE ANOS PARA CONCLUSÃO DE CURSO	93
TABELA 20 - NÚMERO DE DIPLOMADOS.....	93
TABELA 21 TAXAS DE ABANDONO	94
TABELA 22 - RÁCIO ALUNO/PROFESSOR.....	95
TABELA 23 - EVOLUÇÃO BOLSAS DE ESTUDO.....	95
TABELA 24 - NUMERO DE CANDIDATOS REGIME M23	96
TABELA 25 - Nº DE CANDIDATOS, VAGAS POR REGIME DE INGRESSO M23	96
TABELA 26 – VALOR DAS PRESTAÇÕES SERVIÇO	96

LISTA DE FIGURAS

FIGURA 1 - EXEMPLO DE INTEGRAÇÃO.....	11
FIGURA 2 - EXEMPLOS DE DADOS BASEADOS EM APLICAÇÕES/ASSUNTOS.....	12
FIGURA 3 EXEMPLO DE NÃO VOLÁTIL.....	13
FIGURA 4 VARIÁVEL EM RELAÇÃO AO TEMPO.....	13
FIGURA 5 ARQUITECTURA GLOBAL.....	16
FIGURA 6 - ARQUITECTURA DE DATA MART INDEPENDENTE.....	17
FIGURA 7 - ARQUITECTURA DE DATA MART INTEGRADO.....	18
FIGURA 8 ARQUITECTURA DE DATA WAREHOUSE.....	19
FIGURA 9 - IMPLEMENTAÇÃO TOP DOWN.....	21
FIGURA 10 - IMPLEMENTAÇÃO BOTTON UPIMPLEMENTAÇÃO BOTTON UP.....	22
FIGURA 11 - IMPLEMENTAÇÃO COMBINADA.....	23
FIGURA 12 - EXEMPLOS DE GRANULARIDADE.....	25
FIGURA 13 - RELEVÂNCIA DA GRANULARIDADE VERSUS TEMPO.....	26
FIGURA 14 - MODELO ESTRELA.....	30
FIGURA 15 - EXEMPLO DE TABELA DE FACTO.....	31
FIGURA 16 - ESQUEMA ESTRELA.....	34
FIGURA 17 - ESQUEMA EM FLOCO DE NEVE.....	35
FIGURA 18 - MODELO EM CONSTELAÇÃO – KIMBAL.....	36
FIGURA 19 - EXEMPLO DE DATA MART.....	37
FIGURA 20 - MODELO DE IMPLEMENTAÇÃO.....	40
FIGURA 21 - ETAPAS DA METODOLOGIA.....	45
FIGURA 22 - METODOLOGIA DE IMPLEMENTAÇÃO <i>MICROSOFT</i>	50
FIGURA 23 - DIAGRAMA DO SISTEMA DE ENSINO SUPERIOR PORTUGUÊS.....	57
FIGURA 24 - ÁREAS E EIXOS DE INTERVENÇÃO.....	64
FIGURA 25 - OBJECTIVOS E SÍNTESE DAS LINHAS DE ACÇÃO.....	65
FIGURA 26 - COMPONENTES DO DW.....	72
FIGURA 27 - ACTIVIDADES DO DW.....	73
FIGURA 28 - ESQUEMA PARA LEVANTAMENTO REQUISITOS E ANÁLISE E MODELAÇÃO.....	80
FIGURA 29 - ESQUEMA ESTRELA MATRICULAS ALUNOS.....	83
FIGURA 30 - ESQUEMA ESTRELA FASES.....	84
FIGURA 31 - ESQUEMA ESTRELA CANDIDATURAS.....	85
FIGURA 32 - ESQUEMA ESTRELA CONTA ALUNOS.....	87
FIGURA 33 - ESQUEMA ESTRELA IDADES ALUNOS.....	89
FIGURA 34 - ESQUEMA ESTRELA NACIONALIDADES ALUNOS.....	90

LISTA DE ABREVIATURAS

BD – Base de dados

BI – *Business Intelligence*

DAS – Área de tratamento de dados

DBA – Administrador de base de dados

DBMS – *Data Base Management System*

DM'S – *Data Marts*

DSS – Sistemas de suporte à decisão

DW – *Data Warehouse*

EIS – Sistemas de suporte executivo

ER – Entidade e relacionamento

ETL – Extração, transformação e carga

MML – *Multidimensional Modeling Language*

ODS – *Operational Data Store*

OLAP – *OnLine Analytical Processing*

OLTP – *OnLine Transaccional Processing*

SGBD – Sistema de gestão de bases de dados

SI – Sistema de informação

TIC – Tecnologias de informação e comunicação

UML – *Unified Modeling Language*

GLOSSÁRIO

Ad-hoc - São as consultas geradas pelo próprio utilizador de acordo com suas necessidades de cruzar as informações de uma forma não vista e com métodos que o levem a descoberta daquilo que procura.

Agregações - Linhas físicas numa base de dados, criadas, na maioria das vezes, pela soma de outros registos visando melhorar o desempenho de consulta.

Área de Estágio - *Data Staging Area*. Área de armazenamento temporário e conjunto de processos que limpam, transformam, combinam, eliminam duplicidades, armazenam e preparam dados de origem para serem usados no *Data Warehouse*.

Atributo - Coluna (campo) duma tabela da base de dados.

Base de dados multidimensional - Base de dados em que os dados são apresentados na forma de cubos de dados.

Chave composta - Numa base de dados é uma chave formada por várias colunas ou atributos duma tabela.

Chave estrangeira (FK) - Coluna ou atributo numa tabela de base de dados relacional, cujos valores são extraídos de valores de uma chave primária localizada noutra tabela.

Chave primária (PK) - Coluna ou atributo numa tabela de base de dados que é exclusivamente diferente de cada linha na tabela.

Coluna - Estrutura de dados que contém um item de dados específicos dentro de uma linha (registro). Equivalente a um campo de base de dados.

Consulta (*Query*) - Solicitação feita pelo utilizador de informações armazenadas num *Data Warehouse*.

Data Warehouse - Aglomeração das áreas de *staging* e de apresentação do *Data Warehouse* de uma empresa, em que dados operacionais estão estruturados especificamente para consulta e desempenho de análise e facilidade de utilização.

Desnormalização - Aceitação de redundância numa tabela para que ela permaneça simples e não normalizada. Esse procedimento tem por objectivo melhorar o desempenho e facilitar ainda mais a utilização.

Dimensão - Entidade independente num modelo dimensional, que serve como ponto de entrada ou como mecanismo para aplicar o recurso de “separação e combinação” (*slicing and dicing*) das medidas aditivas localizadas na tabela de factos do modelo dimensional.

Drill across, (interligação) - Acto de solicitar dados com a mesma identificação de duas ou mais tabelas de factos num único relatório, quase sempre, envolvendo consultas separadas que são intercaladas com uma segunda passagem através da correspondência de cabeçalhos de linhas.

Drill down (descer na hierarquia) - Acto de adicionar um cabeçalho de coluna ou substituir um cabeçalho de linha por um relatório para dividir as linhas pelo conjunto de respostas mais minuciosamente.

DSS – Sistemas de Suporte à Decisão.

Escalabilidade - Capacidade de acomodar exigências de crescimento futuras.

Extracção, Transformação e Carga (ETL) - Conjunto de processos através dos quais os dados operacionais de origem são preparados, transformados e carregados para o *Data Warehouse*.

Ferramenta de acesso a dados - uma ferramenta cliente que consulta, obtém ou manipula dados armazenados numa base de dados relacional, de preferência, um modelo dimensional localizado na área de apresentação de dados.

Ferramenta de ETL - Aplicação de software normalmente residente no cliente e no servidor que ajuda nos processos de extracção/transacção da carga de dados da produção.

Granularidade - Nível de detalhe capturado no *Data Warehouse*.

Inscrição - O acto pelo qual o aluno, tendo matrícula válida na escola, fica em condições de frequentar as diversas disciplinas em que se inscreve.

Índice - Estrutura de dados associada a uma tabela que esta ordenada logicamente pelos valores duma chave, que é usada para melhorar o desempenho da base de dados e velocidade de acesso a consultas.

Matrícula - o acto pelo qual o aluno dá entrada numa determinada escola, independentemente de, no ano lectivo anterior, ter ou não frequentado um outro estabelecimento de ensino superior. A Matrícula, por si só, não dá direito à frequência, sendo necessário proceder à inscrição anual nas disciplinas do respectivo curso.

Metadados - Qualquer dado mantido para sustentar as operações ou a utilização de um *Data Warehouse* e que funciona como uma enciclopédia para o *Data Warehouse*.

Modelação dimensional - Metodologia que permite modelar logicamente dados para melhorar o desempenho de consultas e disponibilizar facilidade de utilização a partir do conjunto de eventos básicos de medição.

Normalização - Técnica de modelação lógica que remove redundância de dados separando os dados em várias entidades distintas, cada uma das quais se torna uma tabela num SGBD relacional.

OLAP (Processamento analítico *online*) - É um conjunto de princípios vagamente definidos, que fornece uma estrutura dimensional de suporte à tomada de decisões.

OLTP (Processamento de transacções *online*) - Descrição original de todas as actividades e sistemas associados à entrada segura de dados numa base de dados.

Processamento analítico - Utilização de dados, para fins de análise, que visam suportar a tomada de decisões na empresa.

SGBD (Sistema Gestão de Base de Dados) - É um sistema para armazenar, recuperar e modificar dados numa forma altamente estruturada.

Sistema de suporte à tomada de decisões (*DSS*, *Decision Support System*) - Nome original do processo de *Data Warehouse*.

Slice and Dice - Capacidade de aceder igualmente um *Data Warehouse* através de qualquer uma das suas dimensões. É o processo de separação e combinação de dados do *warehouse* em combinações aparentemente infinitas.

Snowflake (Floco de neve) - Dimensão normalizada em que uma dimensão de tabela simples é decomposta numa estrutura de árvore com, potencialmente, muitos níveis de alinhamento.

SQL (*Structured Query Language*) - Linguagem padrão para acesso a bases de dados relacionais.

Tabela - Colecção de linhas (registos) que possuem colunas associadas (campos).

Tabela de dimensão - Tabela num modelo dimensional com uma chave primária composta por uma única parte e colunas de atributos descritivos.

Tabela de factos sem factos - Tabela de factos que não possui factos, mas captura determinados relacionamentos de muitos para muitos entre as chaves de dimensão. É normalmente usada para representar eventos ou fornecer informações que não aparecem nas outras tabelas de factos.

1. INTRODUÇÃO

A evolução das Tecnologias de Informação e Comunicação (TIC) é de tal forma contundente que vai potenciando a descoberta de novos métodos de trabalho, novas formas de comunicação e de descoberta de informação, reduzindo o custo de algumas actividades e permitindo a descoberta de novas áreas de negócio.

Apesar de sermos confrontados com novidades a cada dia que passa no âmbito das tecnologias, os velhos problemas continuam a acontecer relativamente à falta de acesso à informação em tempo útil, à informação de qualidade, bem como na usabilidade das aplicações à nossa disposição.

O uso adequado das tecnologias e mais concretamente da informação permite às organizações não só retirar vantagens sobre as suas concorrentes, assim como, obter um maior conhecimento das suas necessidades e desenvolver iniciativas eficazes visando orientar os seus esforços na satisfação das suas verdadeiras necessidades.

Os sistemas de gestão de informação utilizados pela maioria das organizações, na nossa sociedade, são construídos com suporte em sistemas de bases de dados relacionais, que são pela sua natureza estrutural, vocacionados para armazenar os resultados das operações do dia-a-dia das organizações revelando altos níveis de eficiência.

A proliferação deste tipo de aplicações informáticas ou de sistemas transaccionais tem originado um crescente acumular de dados, o que não significa necessariamente um aumento de conhecimento sobre o negócio. Uma das razões para isto acontecer, relaciona-se muitas vezes com a falta de um modelo de dados global da organização, é devido a muitas das aplicações terem sido desenvolvidas de forma autónoma, tendo como consequência o aparecimento da elevada redundância e inconsistência de informação.

Quem tem responsabilidades na tomada de decisão enfrenta cada vez mais dificuldades originadas por vários factores: previsão de problemas que possam surgir no futuro e na concepção de soluções práticas de resolução desses mesmos problemas, de forma a atingir os objectivos da organização. Cada vez mais os gestores necessitam de estar não

só bem informados, mas também munidos da informação certa, pois esta é a base de toda e qualquer tomada de decisão.

A expressão de que a informação é o maior bem de uma organização, torna-se cada dia que passa mais usual e verdadeira. É de realçar que em grande parte, estas informações estão normalmente disponíveis nas próprias organizações. Os dados para a obtenção da informação estão armazenados nas bases de dados operacionais, nos sistemas proprietários, nas folhas de cálculo, nos ficheiros pessoais e nos documentos. O grande desafio é torná-los úteis, transformando dados em informações que visem acções com vista à melhoria da gestão e tomada de decisão.

Como aceder a toda essa informação dispersa, agregá-la num repositório viável e confiável? A resposta é simples: transformar, arrumar e organizar os dados para a criação de informações estratégicas úteis.

1.1. PROBLEMA DE ESTUDO

As Instituições de Ensino Superior (IES), na sua maioria ou talvez na totalidade, possuem uma base de sistemas de informação, que lhes oferece um certo nível, ou porque não dizer, um bom nível de informatização organizacional. Estes sistemas, aquando do seu desenvolvimento, possuíam como objectivos a automatização de processos e procedimentos que manipulassem grande quantidade de dados, que visam a eliminação de inconsistências e repetições de trabalho. Muitas vezes, estes sistemas foram desenvolvidos sobre tecnologias que não permitem a retirada de uma forma eficaz, rápida e eficiente de informações de carácter estratégico nem a sua eficaz disponibilização.

Perante este cenário, as IES continuam a ter os seus dados em sistemas extremamente robustos, porém pouco flexíveis, o que torna estas organizações mais lentas no acesso à informação e menos preparadas perante os seus concorrentes.

A adopção de tecnologias apropriadas, para o correcto armazenamento e organização de dados, irão permitir às IES viabilizar a sua utilização no melhoramento da sua imagem e na satisfação dos seus clientes (alunos, professores, sociedade). Neste contexto, este projecto pretende mostrar a importância da criação e utilização de um *Data Warehouse*,

que funcionará como suporte da informação no auxílio às tomadas de decisão pelos gestores da organização.

1.2. OBJECTIVOS

A evolução dos Sistemas de Informação, nas últimas décadas, originou um aumento de informação em bases de dados empresariais, possivelmente, instaladas em plataformas heterogéneas, que com o decorrer do tempo tornam o seu manuseamento, a sua análise e inconsistência um problema de proporções preocupantes (Chadhuri, 1997).

Deste modo, é necessário o desenvolvimento de um repositório de dados, otimizados para processos de interrogação mais efectivos, independente dos sistemas operacionais, mas que destes obtivesse a informação desejada. Esta informação é extraída, tratada e guardada no que é chamado de *Data Warehouse* (DW). No entanto, o projecto de desenvolvimento que acede aos dados dos variados sistemas operacionais, que os extrai e transforma de forma a serem inseridos numa base de dados, representando o modelo de dados global da organização é um projecto complexo e moroso.

Propor e implementar um modelo ou metodologia de *Data Warehouse* que contemple o fornecimento de informações precisas e consistentes no âmbito de uma Instituição de Ensino Superior (IES), é o grande objectivo.

Foram definidos os seguintes objectivos específicos:

- Apresentar os benefícios da utilização de um *Data Warehouse* de suporte à tomada de decisão;
- Propor uma metodologia genérica para a implementação de um sistema de *Data Warehouse* de suporte ao sistema de informação;
- Identificar os processos de fornecimento de informações numa IES, listando os aspectos que contribuem para a elaboração do modelo de *Data Warehouse*;
- Identificar os principais problemas que surgem com a extracção de dados dos actuais sistemas de informação, nas IES;
- Modelar, implementar e testar um protótipo numa IES com a finalidade de validar os desenvolvimentos realizados neste trabalho;
- Divulgar à sociedade o resultado deste projecto.

Uma vez definido o projecto de desenvolvimento do DW, deverá ser feita uma análise crítica às propostas existentes de autores conceituados na área de *Data Warehousing*, com particular incidência nas abordagens propostas por William H. Inmon e Ralph Kimball

1.3. METODOLOGIA

O desenvolvimento de um projecto de *Data Warehouse*, similarmente a qualquer outro projecto que envolva tecnologias de informação, deve ter por base a arquitectura da organização em termos de negócio e tecnologias de informação. Essa arquitectura deve estar definida de forma a que, quer as aplicações decorrentes do *Data Warehouse*, quer as outras aplicações estejam alinhadas com o verdadeiro negócio da organização.

A metodologia utilizada na realização do trabalho irá consistir nas seguintes etapas:

- Levantamento teórico sobre as diversas abordagens de criação de um *Data Warehouse*;
- Levantamento teórico sobre desenvolvimento e implementação de um *Data Warehouse*;
- Definição das necessidades de informação nas Instituições de Ensino Superior, em especial no caso particular do Instituto Politécnico de Tomar;
- Desenvolvimento do Protótipo do Sistema de Informação utilizando a arquitectura proposta.

Nas primeiras duas fases do trabalho, aprofundar-se-á o conhecimento teórico e o tipo de abordagens sobre a criação, o desenvolvimento e a implementação de sistemas de *Data Warehouse*. O ênfase concentrar-se-á nas metodologias, nas aplicações, nas necessidades e nas dificuldades quer na implementação quer no desenvolvimento de sistemas de informação, de ambientes de *Data Warehousing* e nas características do ambiente ensino superior.

As últimas etapas do trabalho centrar-se-ão no levantamento das necessidades de informação do caso específico do Gabinete de Avaliação e Qualidade do Instituto Politécnico de Tomar, e no desenvolvimento de um modelo de repositório de dados, que dê suporte aos processos de avaliação dos cursos ministrados no I.P.T. e exigidos pela A3es Agência de e Avaliação e Acreditação do Ensino Superior, e recomendados pela Associação Europeia de Universidades (*European University Association*).

1.4. ESTRUTURA DO TRABALHO

Este projecto reflecte o percurso do trabalho efectuado para o cumprimento do seu objectivo. Deste modo este trabalho foi estruturado em seis capítulos, distribuídos segundo a sequência apresentada a seguir:

O primeiro capítulo expõe uma visão geral do trabalho: os objectivos, a formulação e relevância do problema de estudo, a metodologia a seguir e a organização do estudo.

O segundo capítulo apresenta o resultado de uma revisão de literatura efectuada sobre as áreas relevantes para os objectivos do trabalho, nomeadamente, os conceitos relacionados com *Data Warehouse*, o que permitiu clarificar alguns conceitos teóricos fundamentais para a elaboração deste trabalho.

O terceiro capítulo descreve um conjunto de metodologias de projecto de desenvolvimento de um *Data Warehouse*.

O quarto capítulo apresenta alguns conceitos, inerentes ao Ensino superior em Portugal, duas das entidades que avaliam a qualidade do ensino superior e descreve a organização do Instituto Politécnico de Tomar, nomeadamente, o Gabinete de Avaliação e Qualidade onde é efectuado internamente o processo de avaliação dos cursos.

O quinto capítulo, apresenta o desenho metodológico a seguir no caso de estudo do Instituto Politécnico de Tomar.

O sexto capítulo finaliza com as considerações finais relativas ao projecto identificando algumas recomendações para trabalhos futuros.

2. ENQUADRAMENTO TEÓRICO

2.1. DATA WAREHOUSE

Este sub-capítulo descreve o ambiente de um *Data Warehouse*, os seus conceitos, características, tipos de implementação e granularidade.

2.1.1 CONCEITOS BÁSICOS

Para definir o conceito de *Data Warehouse* (DW) ou armazém de dados é necessário, em primeiro lugar, esclarecer alguns conceitos/termos importantes para a sua compreensão (Sezões e Baptista, 2006, p. 32).

O primeiro é o termo «dados» sendo no DW compostos por factos¹ individuais e discretos que congregam valores descritivos, qualitativos e quantitativos de interesse e relevância para o negócio da organização. No domínio do DW, podemos identificar dois tipos de dados:

- os *operacionais*, que descrevem as transacções e os acontecimentos diários.
- os *informativos*, que são reconciliados, integrados, «limpos» e que constituem a matéria-prima a partir da qual a informação é construída.

O segundo termo é «informação» definida como uma colecção de dados organizada e apresentada num contexto específico e com significado para o negócio. A informação é composta por dados que fazem a “diferença” ou que adicionam valor. O objectivo da informação é informar pessoas, fornecer factos e métricas vitais para os processos de tomada de decisão.

À informação, são vulgarmente atribuídas as características de propósito, de significado, de comunicação, de informar, de recursos e de sentido sintáctico.

*“L’information est un objet formatté (dote de formes reconnaissables)
artificiellement créé par l’être humain à fin de représentation d’un type*

¹ Factos, são uma colecção de itens de dados relacionados, uma medida ou uma leitura de um acontecimento, originado por um determinado processo

d'événement perceptible et identifiable par lui dans le monde réel” (Le Moigne, 1978, p. 29).

Como refere (Gouveia, 2004), informação é um agrupamento de dados, que, relacionados e inseridos num contexto útil, e com sentido bem definido, reduzem a incerteza na tomada de decisão, e influem decisivamente na acção.

Finalmente, o *Data Warehouse* é um repositório de dados, que funciona como uma fonte única e integrada de informação de suporte à decisão e que é formada pela recolha de dados de múltiplas fontes, internas ou externas à organização. Na tabela 1, foram recolhidas algumas definições de autores que passo a mencionar.

Autor	Definição
(Inmon W. , 2002)	<i>Data Warehouse</i> é uma colecção de dados, orientados ao assunto, integrados, variáveis com o tempo e não voláteis. Trata-se de um processo em andamento, que dão suporte ao processo de tomada de decisão, e que aglutina dados de fontes heterogêneas, incluindo dados históricos.
(Harjinder e RAO, 1996)	<i>Data Warehouse</i> é um processo em andamento que aglutina dados de fontes heterogêneas, incluindo dados históricos e dados externos para responder às necessidades de consultas estruturadas, consultas <i>ad-hoc</i> , relatórios analíticos e de suporte à decisão.
(Barquini, 1996)	<i>Data Warehouse</i> é uma colecção de técnicas e tecnologias que conjuntamente disponibilizam um enfoque pragmático e sistemático para tratar com o problema do utilizador final em aceder a informações que estão distribuídas pelos vários sistemas da organização.
(Devlin, 1997)	<i>Data Warehouse</i> é um armazém consistente e completo de dados obtidos de diferentes origens.
(Kimball, L., M., & W., 1998)	<i>Data Warehouse</i> é uma fonte de dados consultáveis da organização, formado pela união de todos os <i>Data Marts</i> correspondentes.

Tabela 1 - Definições de data warehouse

Como se pode verificar, o *Data Warehouse* é um armazém de dados integrados, orientados por assuntos, variáveis no tempo, não voláteis e que disponibiliza a base estrutural para as aplicações de apoio à decisão usadas na organização.

Característica	Definição
integrados	O <i>Data Warehouse</i> deverá ser uma fonte única e abrangente de informação para e sobre o negócio. Assim, com um DW não é necessário aceder a múltiplas fontes de dados para responder a questões levantadas pelos utilizadores;
orientados por assuntos	Os dados são organizados em torno de assunto devendo as informações ser apresentadas compartimentadas por assuntos, de acordo com as necessidades dos utilizadores finais;
variam no tempo	O <i>Data Warehouse</i> contém o histórico da informação, assim como a informação actual, sobre o negócio, possibilitando desta forma análises de evolução histórica;
não são voláteis	<i>Data Warehouse</i> contém informação estável, isto porque os dados, uma vez chegados ao DW, não são apagados. Sendo os dados sempre adicionados de forma regular, de modo a fornecer registos históricos correctos e credíveis; No fundo um repositório fisicamente separado das aplicações operacionais e sem actualizações directas de dados por parte destas.
são acessíveis	Uma das funções primárias de um <i>Data Warehouse</i> é proporcionar o acesso à informação de uma forma fácil e rápida.

Tabela 2 – Características do data warehouse

O *Data Warehouse* é desenvolvido para responder às necessidades informacionais da organização, de forma a sustentar a tomada de decisão, transformando os dados de

várias fontes em informações relevantes para o negócio. Desta forma, a criação de um *Data Warehouse* justifica-se essencialmente por dois motivos:

- A necessidade de integração de dados distribuídos nas diferentes estruturas de bases de dados, tendo em vista uma análise global;
- A necessidade de separação dos dados por grandes assuntos.

Um *Data Warehouse* agrupa frequentemente um vasto manancial de informação, que é muitas vezes dividida por assuntos em conjuntos mais pequenos e agrupados de forma lógica em pequenas unidades designadas como *Data Marts*.

Existem várias ideias, opiniões, definições, conceitos e ferramentas para a criação e manutenção de um *Data Warehouse*, mas é importante ressaltar que não existe nenhum *Data Warehouse*, pronto a ser utilizado sem um trabalho anterior de levantamento de necessidades da empresa.

Segundo (Inmon W. H., 1997), “Os *Data Warehouses* não são construídos de uma só vez. Em vez disso, eles são projectados e povoados passo a passo, portanto, evolucionários e revolucionários.”

2.1.2 CARACTERÍSTICAS

As principais características de um DW são: ser integrado, ser orientado ou organizado por assunto, não ser volátil e ser variável em relação ao tempo.

2.1.2.1 Integrado

Actualmente ainda é comum encontrar organizações com múltiplas bases de dados por cada sistema de informação. Além disso, muitos desses sistemas na sua construção não tiveram padrões comuns, o que resulta em dados representados de maneiras diferentes. Cada analista de sistemas define a mesma estrutura de dados de várias formas, ocasionando que os dados que representam a mesma informação, sejam representados de diversas maneiras dentro dos sistemas utilizados pela organização ao longo do tempo. Desta forma, na maioria das vezes, é necessário integrar dados oriundos de todos

estes sistemas existentes. Um pequeno exemplo desse problema é a representação do estado civil na Figura 1.

Uma base de dados pode conter um campo do tipo alfanumérico para definir o estado civil. Este campo é definido da seguinte forma: C = “Casado”; S = “Solteiro”; D = “Divorciado”; V = “Viúvo” e O = “outros”. Noutra base de dados a mesma informação pode ser representada por: 1 = “Casado”; 2 = “Solteiro”; 3 = “Divorciado”; 4 = “Viúvo” e 9 = “outros”. Com a integração dos dados, o problema fica resolvido, pois deve ser utilizada uma única representação para esta informação, conforme ilustra a Figura 1.



Figura 1 - Exemplo de Integração (Fonte: Adaptado de Inmon (1997,37))

Como o objectivo de um sistema de *Data Warehouse* (DW), consiste em agregar informação proveniente de uma ou mais Bases de Dados (BD), ou de outras fontes, para posteriormente tratar, formatar e consolidar a informação numa única estrutura de dados, o nível de integração dos dados é a principal característica dum *Data Warehouse*.

2.1.2.2 Orientado por Assunto

Significa que o *Data Warehouse*, armazena as informações agrupadas por assunto de maior interesse da organização isto é as que são consideradas mais importantes. Pelo contrário, os sistemas transaccionais são orientados a processos desenvolvidos para manter as transacções diárias, isto é os sistemas transaccionais estão voltados para processos e aplicações específicas enquanto os DW's objectivam assuntos.

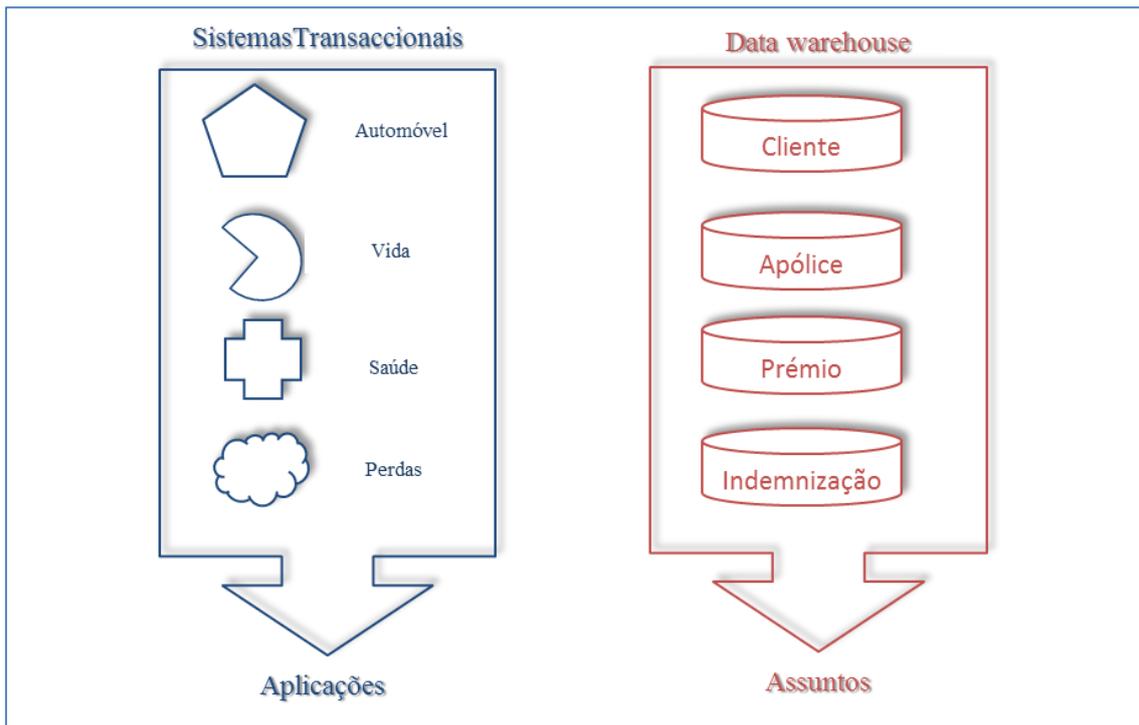


Figura 2 - Exemplos de dados baseados em aplicações/assuntos - Fonte: Adaptado de Inmon (1997)

Ao iniciar a construção dum *Data Warehouse*, devemos discutir com os utilizadores finais quais os seus objectivos, isto é quais são as informações importantes para o processo de análise e no fundo, quais as suas necessidades. Como os dados são organizados em torno de assuntos, as informações finais serão apresentadas compartimentadas por assuntos de acordo com as necessidades dos utilizadores finais.

2.1.2.3 Não volátil

Significa que, num DW existem somente duas operações: a carga dos dados (inicial e incremental) e o acesso a esses dados em modo de leitura (consultas). Isto significa que, os dados não sofrem alterações, são carregados uma única vez e a partir desse momento não são alterados, isto é, apenas podem ser consultados.

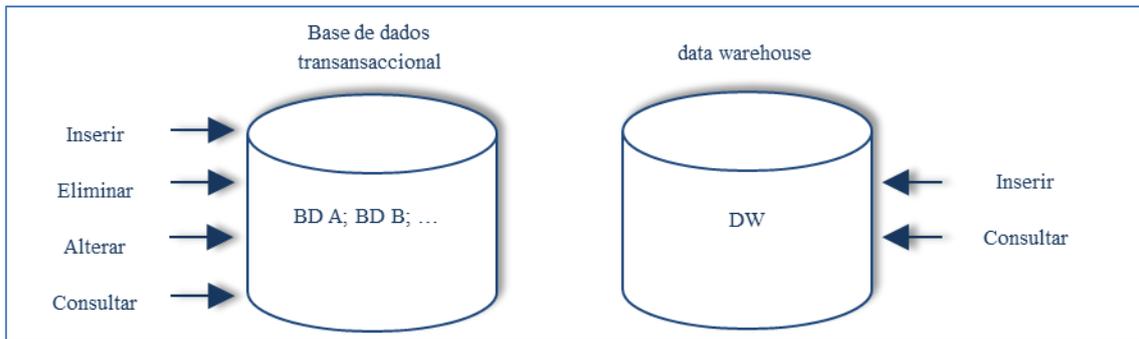


Figura 3 Exemplo de Não Volátil - Fonte: Adaptado de Inmon (1997)

2.1.2.4 Variável em relação ao Tempo

Segundo Inmon, quando se fala em ser variável em relação ao tempo, quer dizer que, o horizonte de tempo para um DW é de anos (5 a 10), isto é, significativamente maior do que nas BD operacionais em que o horizonte de tempo são dias (dados recentes até 60 a 90 dias) (Inmon W. , 2002).

Os dados do DW, são precisos em relação ao tempo, representam resultados operacionais em determinado instante, o momento de tempo em que foram capturados. São um “*snapshot*”, um conjunto estático de registos de uma ou mais tabelas, capturadas num determinado instante predeterminado. Nas BD operacionais os dados podem ser modificados, mas no DW depois de carregados, os dados já não podem ser modificados, são armazenados de forma histórica e apenas fazem sentido se contextualizados com a data da sua criação. Isso implica que os dados num DW não podem ser actualizados.

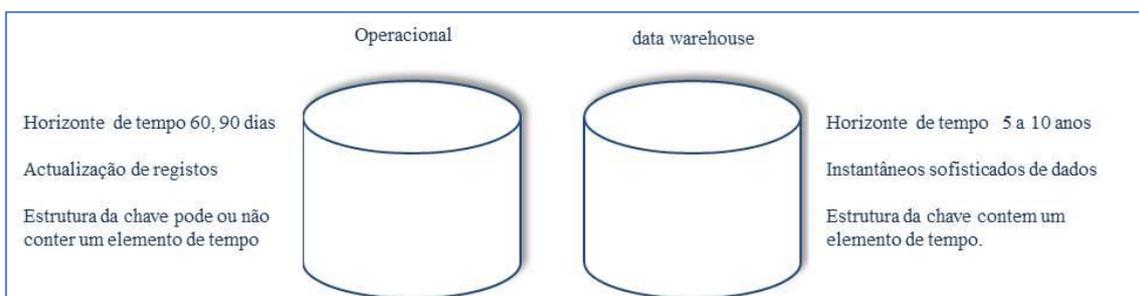


Figura 4 Variável em relação ao tempo - Fonte: Inmon (1997)

2.1.3 ARQUITECTURA DE DATA WAREHOUSE

Uma arquitectura é um conjunto de normas e regras, que estruturam um projecto de um sistema ou produto e que no caso particular do DW auxilia no entendimento do fluxo de dados ao longo do processo de *data warehousing*.

Conforme refere (Machado, 2004, p. 47) “ *A escolha da arquitectura é uma decisão gerencial do projecto, e está normalmente baseada nos fatores relativos à infra-estrutura disponível, ao ambiente de negócios (porte da empresa), concomitantemente com o escopo de abrangência desejado, assim como a capacitação dos empregados da empresa e dos recursos disponibilizados ou projetados para investimento*”.

A escolha da arquitectura é uma decisão que afecta e provoca impactos dramáticos no sucesso de um projecto de *Data Warehouse*, pois existem variáveis como, por exemplo: o tempo; o retorno do investimento; a satisfação dos utilizadores finais que condicionam a a sua escolha.

A literatura fornece duas grandes arquitecturas para *Data Warehouse*, a Arquitectura Global e a Arquitectura de *Data Mart*.

2.1.3.1 Arquitectura Global

Nesta arquitectura global ou de *Data Warehouse* empresarial ou normalizado, o objectivo é a construção do armazém de dados como uma aplicação global, que visualiza o *Data Warehouse* como o centro do ambiente analítico inteiro da empresa como defendido por Inmon,

O *Data Warehouse* é projectado e construído com base nas necessidades da organização como um todo.

Global = Por toda a organização

O primeiro passo corresponde à extracção, transformação, migração e carregamento de dados oriundos dos sistemas operacionais ou de fontes externas. No processo de extracção, os dados são retirados das suas origens e armazenados na área de estágio (*Staging Area*). No processo de transformação, os dados são filtrados, eliminando-se

aqueles que não são necessários e inicia-se a sua transformação de encontro com as necessidades dos requisitos levantados para o projecto de DW. De seguida, no processo de migração, os dados e os metadados necessários são carregados para dentro do *Data Warehouse*.

Uma vez constituído o DW, os *Data Marts* são oriundos a partir de resumos do DW e dos metadados.

Vantagens	Desvantagens
Visão da empresa como um todo	Alta taxa de riscos, por não existir garantias de resultados neste tipo de ambiente.
Herança pelos datamart's da arquitectura do <i>Data Warehouse</i>	Implementação demasiado longa por envolver todos os sectores da organização
Agregação dos dados e medidas	Arquitectura de base ser complexa e, conseqüentemente de difícil compreensão pelos utilizadores normais e não especialistas.
Regras de extracção, limpeza e integração centralizados, facilitando a manutenção.	A demora na construção do <i>Data Warehouse</i> e a falta de retorno rápido, causa ansiedade e expectativas nos utilizadores.

Tabela 3-Vantagens/Desvantagens da Arquitectura Global

Esta arquitectura habilita os utilizadores finais a utilizarem visões corporativas de dados, contudo consome muito tempo de desenvolvimento, administração, e ainda elevados custos de implementação.

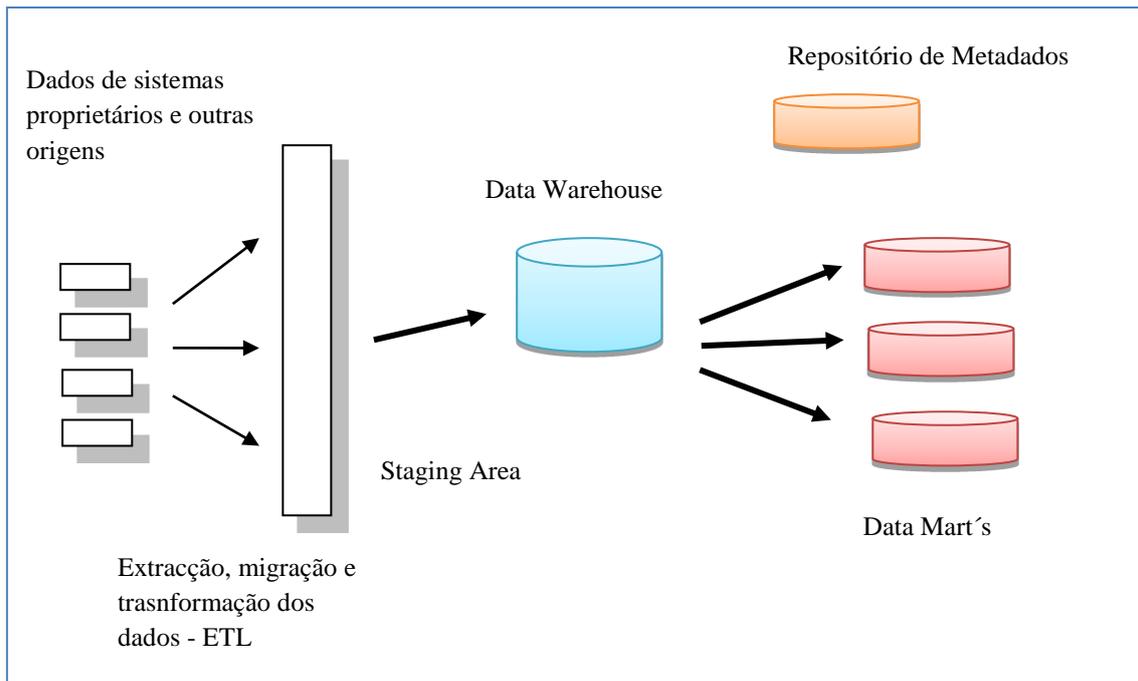


Figura 5 Arquitectura Global - Fonte: Machado (2004,)

2.1.3.2 Arquitectura de *Data Mart*

Esta arquitectura é apresentada por Kimball [1998] e é a preferida dos fornecedores de software porque tem uma implementação rápida, e como característica principal, o processo de desenvolvimento inverso ao proposto por Inmon.

O desenvolvimento do DW é feito de uma forma progressiva isto é, em primeiro lugar são criados os *Data Marts*, a partir dos dados de sistemas fontes ou dos dados externos à organização e posteriormente é que é criado o DW.

Nesta arquitectura o *Data Warehouse* é construído de forma incremental, e os *Data Mart's* são construídos de forma independente. O processo tem início com a construção de um ou mais *Data Marts* que é resultado das necessidades de cada departamento ou área da organização.

Vantagens	Desvantagens
Implementação direccionada, permitindo um desenvolvimento rápido.	Perigo dos <i>Data Mart's</i> se tornarem DM independentes, o que dificultam e por vezes inviabilizam a integração com o DW.
Retorno rápido para os utilizadores	Inexistência de foco corporativo, pois são baseados em datamart's isolados.
Herança das informações pelo DW de forma incremental, reduzindo riscos no projecto.	A maldição do sucesso, começa a perseguir os responsáveis do projecto. Enquanto uns utilizadores estão satisfeitos por terem o seu DM actualizado, outros ainda estão a aguardar queo seu DM seja actualizado

Tabela 4 Vantagens/Desvantagens da Arquitectura *Data Mart*

A Arquitectura de *Data Mart*, pode ser dividida em duas sub-arquitecturas, uma de *Data Mart's* Independentes e outra de *Data Marts* Integrados, conforme os *Data Mart's* estão ou interligados.

a1) Arquitectura de *Data Mart's* independentes, é a arquitectura preferida dos fornecedores de software para consulta de informações no *Data Warehouse*, pois sendo de mais fácil implementação leva a resultados mais rápidos. Esta arquitectura implica *Data Mart's* Stand alone, controlados por um grupo específico de utilizadores e que responde apenas às suas necessidades específicas, sem nenhum foco corporativo.

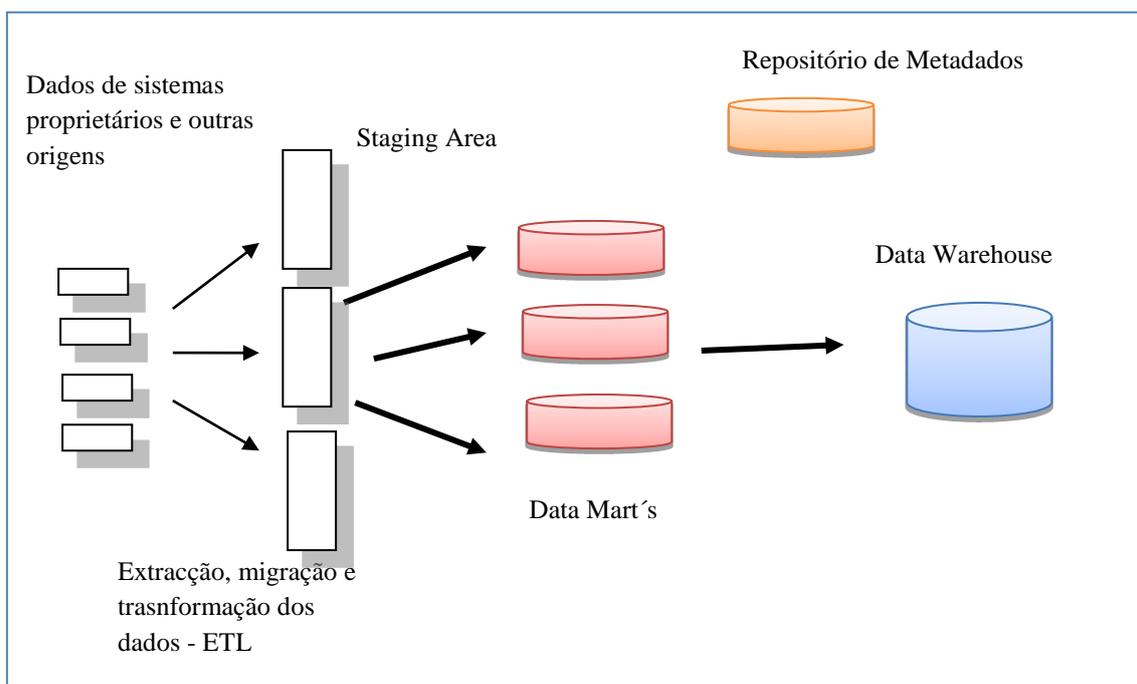


Figura 6 - Arquitectura de Data Mart Independente - Fonte: Machado (2004)

a2) Arquitectura de *Data Marts* Integrados, Combinada ou BUS

Caracteriza-se pelo desenvolvimento de *Data Marts* separadamente por grupos de trabalho, departamentos ou áreas, mas planeados e integrados através de metadados e tabelas de factos e dimensões conformes, sendo possível aceder e utilizar os dados de um DM de outro departamento. Nela, uma única área de extracção é partilhada por todos os *Data Marts*. Este tipo de arquitectura é basicamente uma distribuição de implementação. Apesar dos *Data Marts* serem criados separadamente por grupos de trabalho ou departamento, eles são integrados e interligados, dando assim uma visão corporativa dos dados. Contudo esta integração aumenta sensivelmente o grau de complexidade dos requisitos.

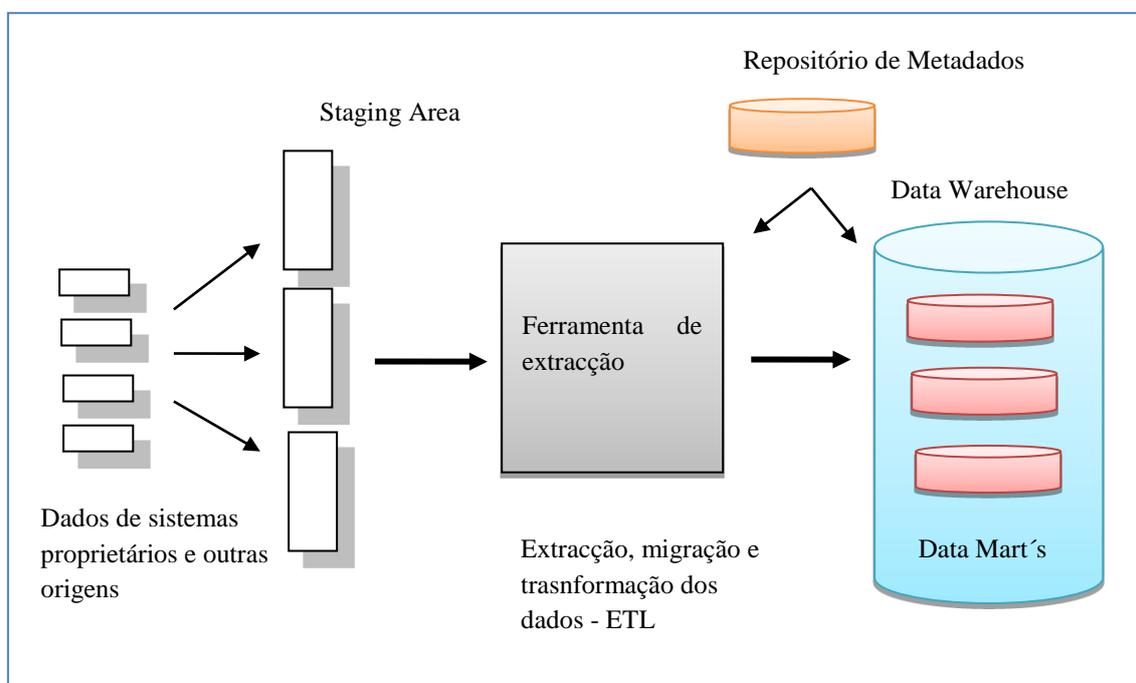


Figura 7 - Arquitectura de Data Mart Integrado - Fonte: Machado (2004)

2.1.3.3 Considerações sobre Arquitectura

A arquitectura de desenvolvimento é um dos aspectos mais importantes na construção de um *Data Warehouse* e poder ser representada pela figura 8.

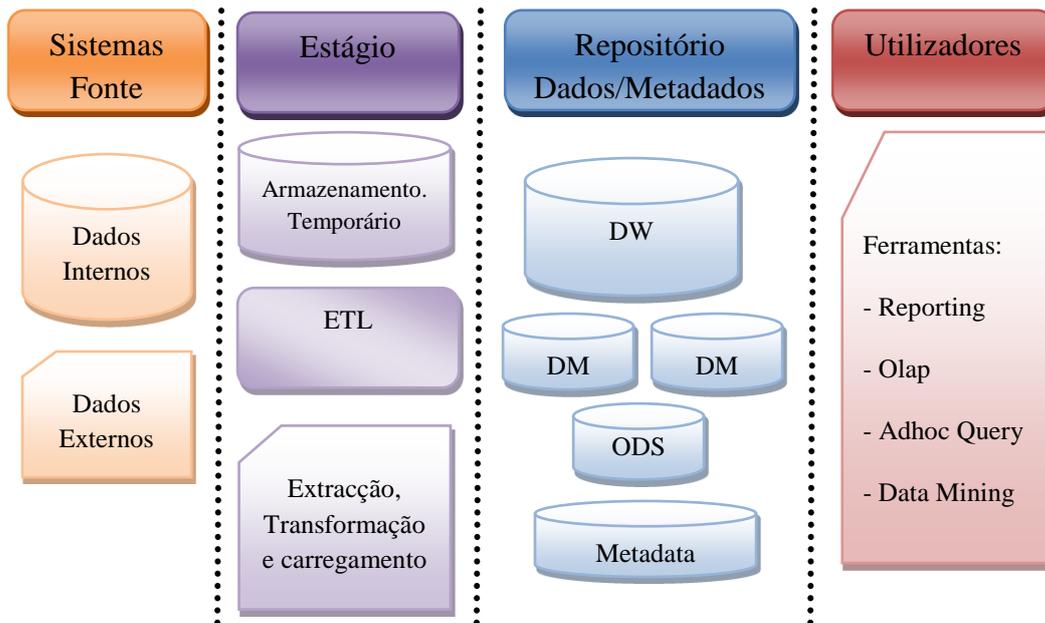


Figura 8 Arquitectura de Data Warehouse – Fonte: Elaboração própria

Os **sistemas fonte ou sistemas operacionais**, são os locais ou sistemas de onde são recolhidos os dados e, tanto podem ser internos ou externos à organização. As fontes internas são provavelmente as mais frequentes e são aquelas em que a organização trabalha no dia a dia. As fontes externas são por exemplo as Tabelas de Nacionalidades, Códigos Postais, Códigos oficiais de cursos e que normalmente recorremos ao exterior para a sua obtenção.

A **zona de estágio** é o subsistema de ligação entre o sistema fonte ou sistema operacional e o *Data Warehouse* propriamente dito. Esta componente compreende uma área de armazenamento temporário de dados e um conjunto de processos conhecidos como ETL (Extract-Transform-Load) (Caldeira, 2008).

Esta zona contém ferramentas para a **Extracção, Transformação e Carga (ETL)**, onde é efectuado o armazenamento através de processos que permitem extrair, transformar e carregar os dados fonte para serem utilizados no DW.

O processo de extracção como o nome indica consiste na leitura e extracção dos dados dos sistemas fontes colocando-os ao serviço da área de estágio. Podemos dizer que tem

duas situações distintas: a primeira extracção de dados propriamente dito e a segunda extracções incrementais.

O processo de transformação consiste na limpeza de erros e inconsistências, transformação dos dados num esquema de informação uniforme e coerente e na eliminação de campos considerados inúteis.

O processo de carregamento também conhecido como de refrescamento desenrola-se através de mecanismos e períodos de tempo distintos.

A área temporária de armazenamento dos dados pode ser definida de diferentes formas. Os dados no momento da sua transferência podem ser guardados em estruturas não estruturadas como folhas de cálculo ou em ficheiros de tipo texto, ou, em alternativa, em estruturas estruturadas como uma base de dados (Caldeira, 2008).

O **repositório dos dados e metadados**, zona onde estão armazenados todos os dados do DW.

De acordo com a metodologia dimensional um *Data Warehouse* não é mais do que a união de todos os *Data Marts* consolidados definidos a partir dos processos de negócio da organização (Caldeira, 2008).

Ao serem informação sobre os dados os metadados são uma parte estruturante muito interessante, e um meio essencial para o sucesso das etapas iniciais de desenvolvimento do *Data Warehouse* (Caldeira, 2008).

Finalmente a **zona de utilizadores**, onde existem as ferramentas que permitem disponibilizar de forma fácil e rápida o resultado das consultas ao DW.

A pesquisa como um facilitador dos processos de análise de dados, é o propósito fundamental da existência de uma *Data Warehouse*. As ferramentas de acesso aos dados são assim o componente que tem uma maior proximidade com o utilizador (Caldeira, 2008).

2.1.4 TIPOS DE IMPLEMENTAÇÃO

Vários tipos de implementação das arquitecturas apresentadas podem ser utilizados como por exemplo a *Top down*, a *Bottom Up* ou uma combinação das duas que normalmente é identificada como de Intermédia.

2.1.4.1 Implementação *Top down*

A implementação top-down é conhecida como padrão inicial do conceito de DW (Machado, 2004). Defendida por Inmon visualiza o *Data Warehouse* como o centro do ambiente analítico inteiro da organização.

Preconiza duas etapas em que na primeira se define o esquema de conteúdo de todo o DW (1). O processo inicia-se com a extracção, a transformação e a integração das informações dos sistemas e dados externos para um ODS (*Operational Data Store*) ou até directamente das fontes operacionais, sendo de seguida os dados transferidos para o DW.

A segunda etapa consiste na implementação de *Data Mart's*, de acordo com as características particulares de cada departamento. A partir do DW são extraídos os dados e metadados para os *Data Mart's*.

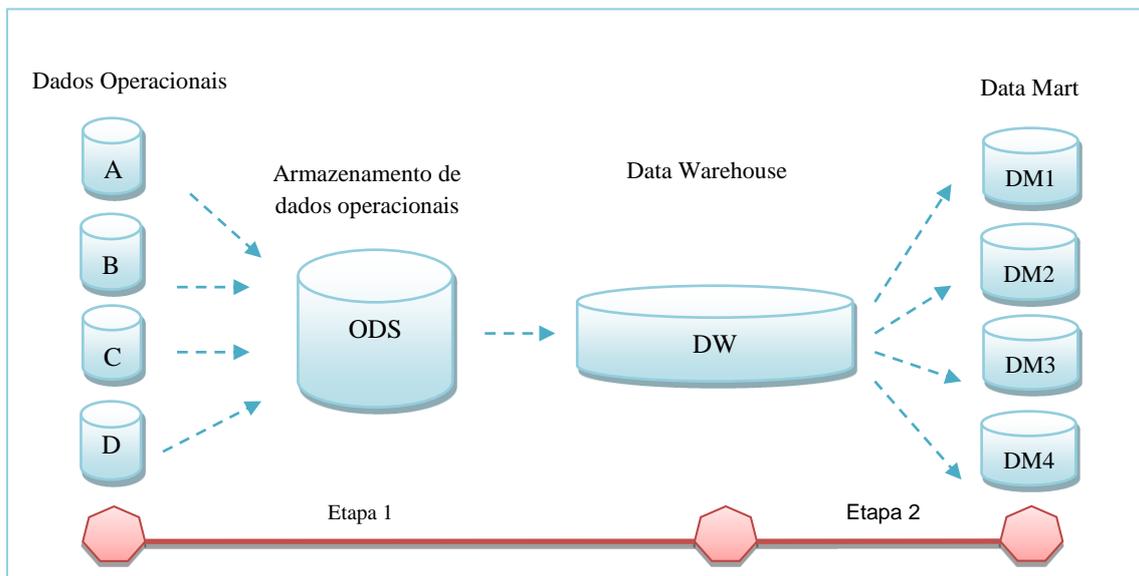


Figura 9 - Implementação Top Down – Fonte: Adaptado de Kimball (1996)

De uma forma geral, quando for seleccionada a implementação *top down* uma arquitectura *Data Warehouse* global será a implementada.

2.1.4.2 Implementação Botton Up

Este tipo de implementação permite que o planeamento e o desenho dos *Data Mart's* possam ser realizados sem esperar que seja definida uma infra-estrutura de corporativa para o DW na organização (Machado, 2004).

Esta abordagem de implementação é defendida por Kimball e tem a sua característica principal um processo de desenvolvimento inverso à proposta de Inmon. Inicialmente são criados os *Data Mart's* a partir dos sistemas operacionais, e na fase posterior será criado o *Data Warehouse*, como a união de todos os *Data Marts's*.

Tem como objectivo modelar e construir esquemas dos conteúdos de cada *Data Mart*, tendo em conta as necessidades informacionais existentes. Os esquemas de cada DM devem ser modelados com o objectivo de posteriormente serem unificados pois, só assim se consegue obter no final um esquema global para todo o DW.

O propósito desta implementação é a construção de um DW incremental a partir de desenvolvimento de *Data Mart's* independentes. O processo inicia-se com a extracção, a transformação e a integração dos dados para um ou mais *Data Mart's*, os quais são modelados, normalmente com base num modelo dimensional.

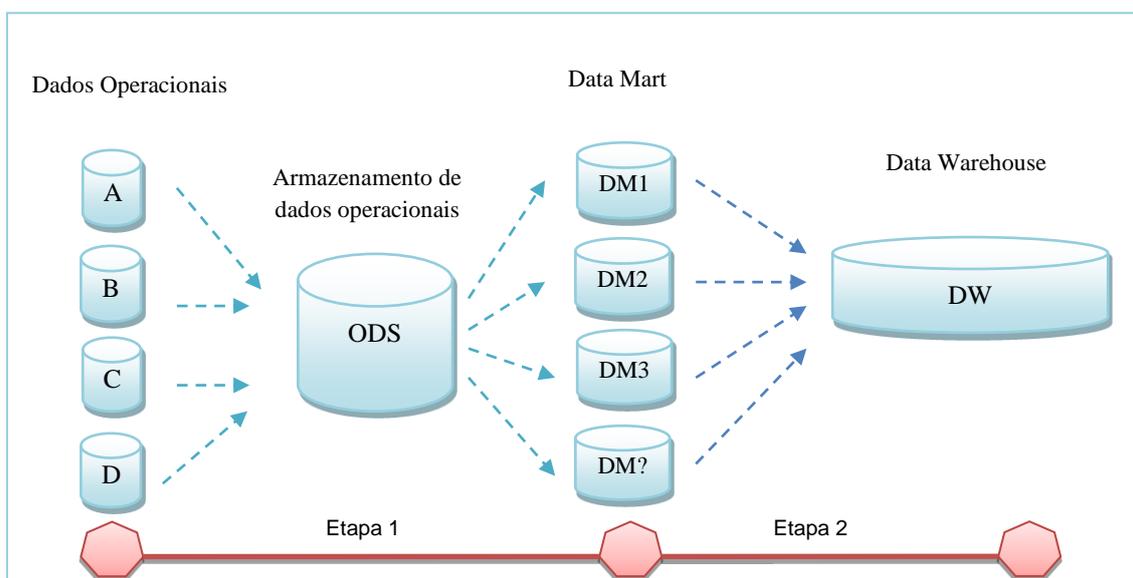


Figura 10 - Implementação Botton Up - Fonte: Adaptado de Kimball (1996)

2.1.4.3 Implementação Combinada

Este tipo de implementação integra as duas anteriores implementações. Nesta abordagem o modelo de dados do DW tem uma visão macro do modelo e no passo seguinte passa-se à implementação de partes desse modelo. Essas partes são processos ou áreas e constituem os DMs. Onde cada um deles pode ser gerado a partir do modelo de dados do DW e integrado ao seu modelo físico (Machado, 2004).

Este tipo de implementação tem como princípios o levantamento de requisitos de forma global mas os *Data Marts* serão desenvolvidos e integrados incrementalmente. A construção do DW inicia-se com o desenvolvimento do primeiro *Data Mart*, definido durante o planeamento. O ciclo repete-se incrementalmente, com a construção dos outros *Data Marts*, sempre respeitando os factos e dimensões comuns, até formarem o *Data Warehouse*. A principal vantagem desta abordagem é a garantia dos dados, em virtude do modelo de dados para os *Data Marts* ser único, possibilitando realizar o mapeamento e o controle dos dados. Como desvantagem, poderão ocorrer complicações políticas por conta da escolha da sequência dos *Data Marts* a serem implementados.

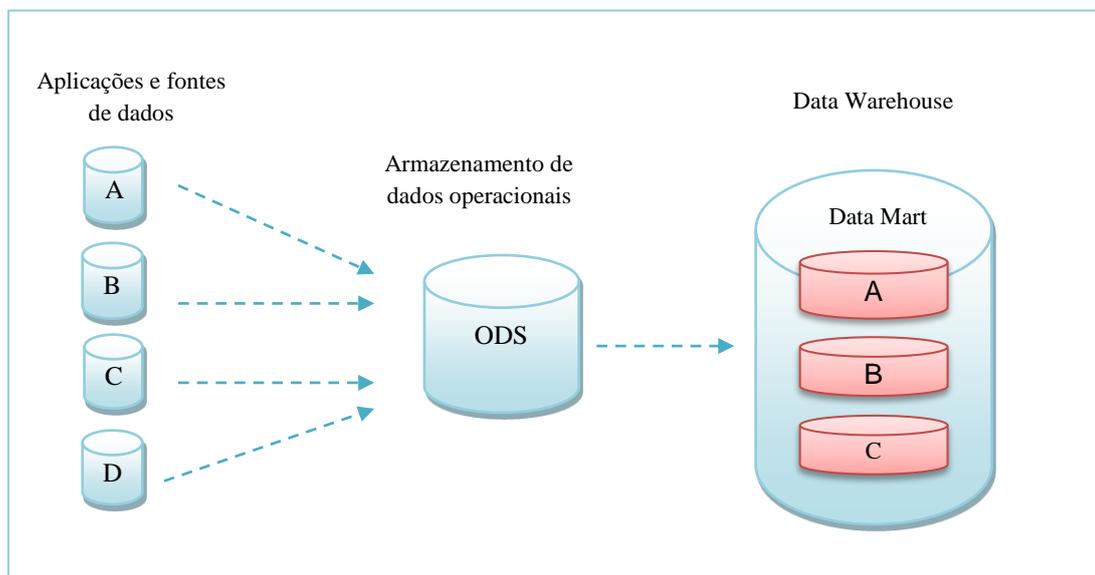


Figura 11 - Implementação combinada - Fonte: Adaptado de Kimball (1996)

2.1.5 GRANULARIDADE

A granularidade diz respeito ao nível de detalhe, sumarização ou resumo dos dados dentro do *Data Warehouse*. Quanto mais detalhe existir, mais baixo será o nível de granularidade; consequentemente, quanto menor for o nível de detalhe, maior será a granularidade. O nível de granularidade afecta directamente o volume de dados armazenados no DW, logo deverá ser considerada como a principal questão dum projecto de DW.

Um nível de granularidade alto irá garantir uma maior rapidez nas consultas, em economia de armazenamento na base de dados. Inversamente um nível de granularidade mais baixo, permite obter respostas a praticamente qualquer consulta, mas haverá a necessidade do aumento de espaço de armazenamento no *Data Warehouse*. Segundo (Inmon W. H., 1996), é necessário o ponto de equilíbrio na escolha dos níveis adequados de granularidade, que deverá estar situado entre o volume de dados para armazenar e a capacidade de responder às necessidades dos utilizadores finais.

A definição do grão é um passo fundamental no processo de desenvolvimento dum *Data Mart*, e é extremamente importante para além de ser rigorosamente definido, que seja determinado com o máximo nível de detalhe, ou o que é a mesma coisa, com a maior atomicidade admissível no ambiente de aplicação. O nível de detalhe do modelo dimensional é essencial na compreensão do negócio em análise, na dimensão do esquema em estrela e nos aspectos técnicos de construção do *Data Warehouse* (Caldeira, 2008).

Inmon define a granularidade como sendo o aspecto mais importante do DW. Essa característica diz respeito ao nível de detalhe ou de resumo das informações que estão no DW (Inmon W. H., 1996).

Muitos dos autores destacam a granularidade apenas como factor de controlo da área de armazenamento de dados e de performance (quanto maior o volume de dados menor a performance), convém no entanto destacar a importância da definição de granularidade num projecto de DW em termos de análise de negócios.

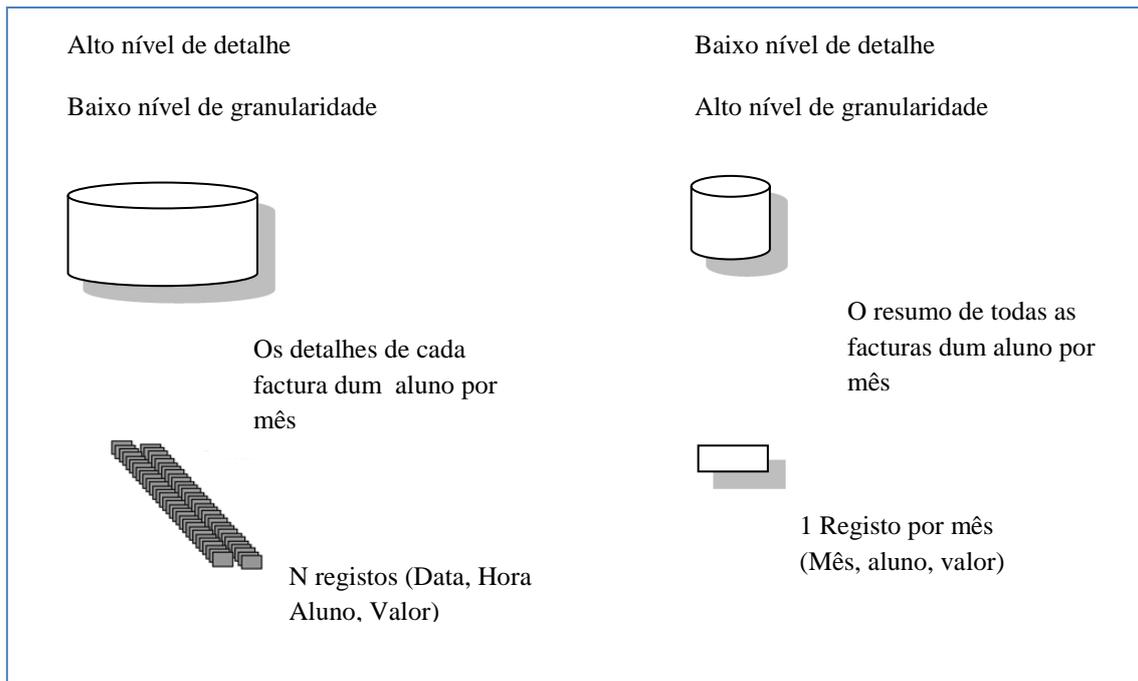


Figura 12 - Exemplos de Granularidade - Fonte: Adaptado de Inmon (1997)

As necessidades actuais e futuras do negócio da organização são o aspecto chave a considerar na escolha do nível de detalhe a aplicar num modelo dimensional. O mínimo que se pode exigir a um grão é que consiga resolver as questões correntes de uma organização. Contudo não devem ser esquecidos que as necessidades futuras do negócio também devem ser acauteladas, pelo que a escolha inicial do nível de detalhe é fundamental para que haja dados disponíveis para usos futuros e, como tal, actualmente desconhecidos. Assim, em caso de dúvidas sobre utilizações futuras do *Data Mart* é boa prática optar sempre pelo grão mais fino possível (Caldeira, 2008).

Aliás o autor Kimball defende também a adopção de baixos níveis de granularidade, pois dessa forma será possível a utilização de ferramentas de *data mining*, que é muito pouco eficiente quando realizada em dados agregados (Kimball, L., M., & W., 1998).

Como forma de solucionar a questão a questão da organização possuir grande quantidade de dados no DW, Inmon propõe dois ou mais níveis de detalhe, para o seu armazenamento:

- Dados de detalhe corrente, enquadram-se aqui todos os dados provenientes dos sistemas fonte que foram inicialmente carregados. A preocupação mais importante em relação aos dados encontra-se no nível de detalhe corrente, pois

reflectem os acontecimentos mais recentes e são considerados de maior interesse para a instituição. Neste nível de detalhe os dados são armazenados em séries diárias com horizonte de tempo de até 2 anos.

- Dados levemente resumidos (o departamental ou *Data Mart*) os dados são resumidos e armazenados neste nível de dados, onde haverá um volume de dados significativamente menor que no anterior. Os dados levemente resumidos compreendem um nível intermédio de detalhe dos dados pelo que o horizonte de tempo utilizado é de até 5 anos e os dados são armazenados em séries semanais e mensais.
- Dados altamente resumidos (o individual) São obtidos a partir do resumo de dados levemente resumidos e tem como finalidade responder aos utilizadores com informações mais compactas. O nível seguinte de dados proposto no projecto de estrutura do DW é o altamente resumidos, onde estão compactos e de fácil acesso. Aqui o horizonte de tempo é de 10 anos e os dados são armazenados em séries anuais ou semestrais.

Também para cada nível de detalhe, é estabelecido um horizonte de tempo, que aumenta inversamente ao nível de detalhe.

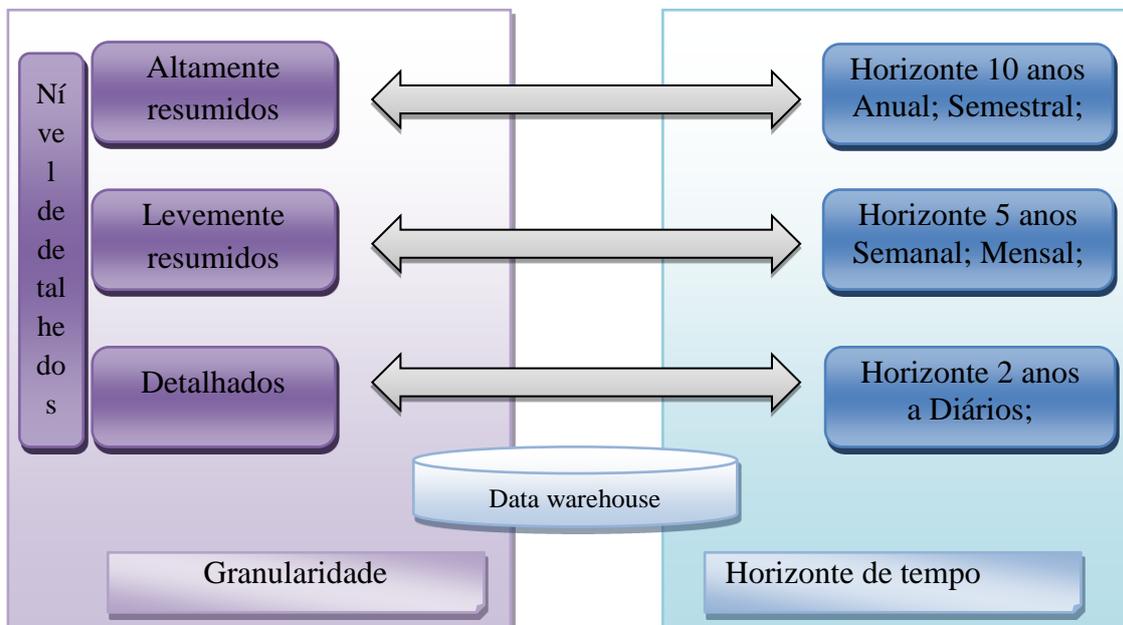


Figura 13 - Relevância da Granularidade versus tempo – Fonte: Elaboração própria

O ponto de equilíbrio na escolha dos níveis adequados de granularidade dá-se entre a gestão do volume de dados e o seu armazenamento, segundo um nível de granularidade tão alto que a utilização detalhada dos dados já não possa ser realizada.

2.2. METADADOS

Metadados são dados sobre os dados. Metadados são todas as informações do ambiente do DW que não são os seus próprios dados (Inmon W. H., 1996).

Ao serem informação sobre os dados, os metadados são uma parte estruturante muito interessante, e um meio essencial para o sucesso das etapas iniciais de desenvolvimento do *Data Warehouse* (Caldeira, 2008).

Enquanto os metadados são em si um tema relativamente recente, os conceitos que lhes estão subjacentes têm estado em uso desde que se fazem compilações de informação. A título de exemplo, os catálogos das livrarias representam um tipo de metadados que há muitas décadas servem como instrumento de gestão de informação.

Sobre a origem da palavra, o termo «meta» exprime a ideia de nível superior, maior generalidade e apresenta-se como um sinónimo de «além de; para além de». Também importante para a obtenção de uma definição de metadados é a aceção corrente do termo «meta» como partícula de constituição de palavras que denota uma lógica reflexiva. Por exemplo, «meta-jornalismo» significa fazer reportagens jornalísticas sobre jornalismo, assim como «meta-fotografia» significa tirar fotos a fotos. Estas duas definições da partícula «meta» estão provavelmente na génese da definição mais frequente da palavra «metadados»: «dados sobre dados» ou «informação sobre dados», definição que peca por alguma falta de precisão.

Os metadados são muitas vezes entendidos como «documentação». Isto mesmo é confirmado por Inmon, ao identificar que no ambiente operacional os metadados são «tratados de forma secundária e relegados para o mesmo nível de importância da documentação» (Inmon W. , 2002).

Podemos então definir metadados como «dados estruturados sobre dados», quando utilizados num contexto em que sejam «associados a um sistema de informação ou a um

objecto informativo», para efeitos de «descrição, administração, imposição de imperativos legais, funcionalidade técnica, utilização» ou manutenção.

2.3. MODELO DE DADOS

A construção do modelo de dados é fundamental para o desenvolvimento, ajudando a compreender as regras de negócio e a própria organização dos dados de forma a otimizar os tempos de resposta (Kimball, 1996).

Este autor aponta também a existência de uma grande diferença no modelo de dados das bases de dados transaccionais e das dimensionais. De facto, os modelos de dados não são apenas diferentes como também têm objectivos diferentes. Enquanto nos sistemas transaccionais o que conta é a velocidade com que se acede e actualiza um dado ou um conjunto de dados (eficiência), num DW, o que conta é a possibilidade e facilidade de dispor os dados agregados com um nível de agregação conveniente (eficácia). Em ambientes transaccionais o modelo de dados mais utilizado é o de Entidade/Relacionamento (ER), enquanto que num ambiente de *Data Warehouse* o modelo mais utilizado é o dimensional.

Desta forma existem os seguintes modelos de dados:

- - Modelo de dados Entidade/Relacionamento (ER);
- - Modelo de dados dimensional.

2.3.1. MODELO DE DADOS ENTIDADE RELACIONAMENTO (ER)

Em 1970, E. F. Codd² publicou um artigo "*Relational Model of Data for Large Shared Data Banks*" com os fundamentos teóricos do modelo relacional. Este modelo utiliza dois conceitos: as entidades e os relacionamentos entre essas entidades. Baseia-se numa estrutura de dados simples e uniforme chamada relação, que pode ser visualizada sobre

² Edgar Frank Codd, Matemático britânico

o formato de uma tabela, onde cada linha representa uma colecção de valores, de atributos, relacionados com uma entidade do mundo real.

Os sistemas OLTP processam diariamente uma quantidade enorme de transacções. Nestes sistemas a consistência transaccional e a velocidade são requisitos imperativos pois operam as actividades essenciais de uma organização (Kimball, L., M., e W., 1998).

O modelo entidade/relacionamento aplica-se perfeitamente a este tipo de sistema, pois divide os dados em várias entidades distintas, cada uma implementada numa tabela na base de dados (Kimball R. , 1998).

2.3.2. MODELO DE DADOS DIMENSIONAL

O modelo dimensional não é uma forma diferente de estruturar um sistema transaccional, pelo contrário o modelo dimensional destina-se apenas a ser um sistema de publicação de dados. Pelo que neste tipo de modelo não se colocam os problemas de integridade dos dados tão frequente no modelo E/R (Caldeira, 2008) .

Ao contrário do modelo entidade/relacionamento, o modelo dimensional é muito assimétrico. Há uma tabela dominante no centro do diagrama com múltiplas ligações a outras tabelas. Cada uma das tabelas secundárias possui apenas uma junção com a tabela central. A tabela central é chamada de tabela de factos. As tabelas secundárias são chamadas de tabelas de dimensão (KIMBALL, 1998).

A figura abaixo mostra um modelo dimensional o qual possui a tabela facto (Conta Candidatos) no centro do diagrama. As dimensões do facto (Cursos, Ano lectivo, Candidatos) aparecem em redor da tabela central.

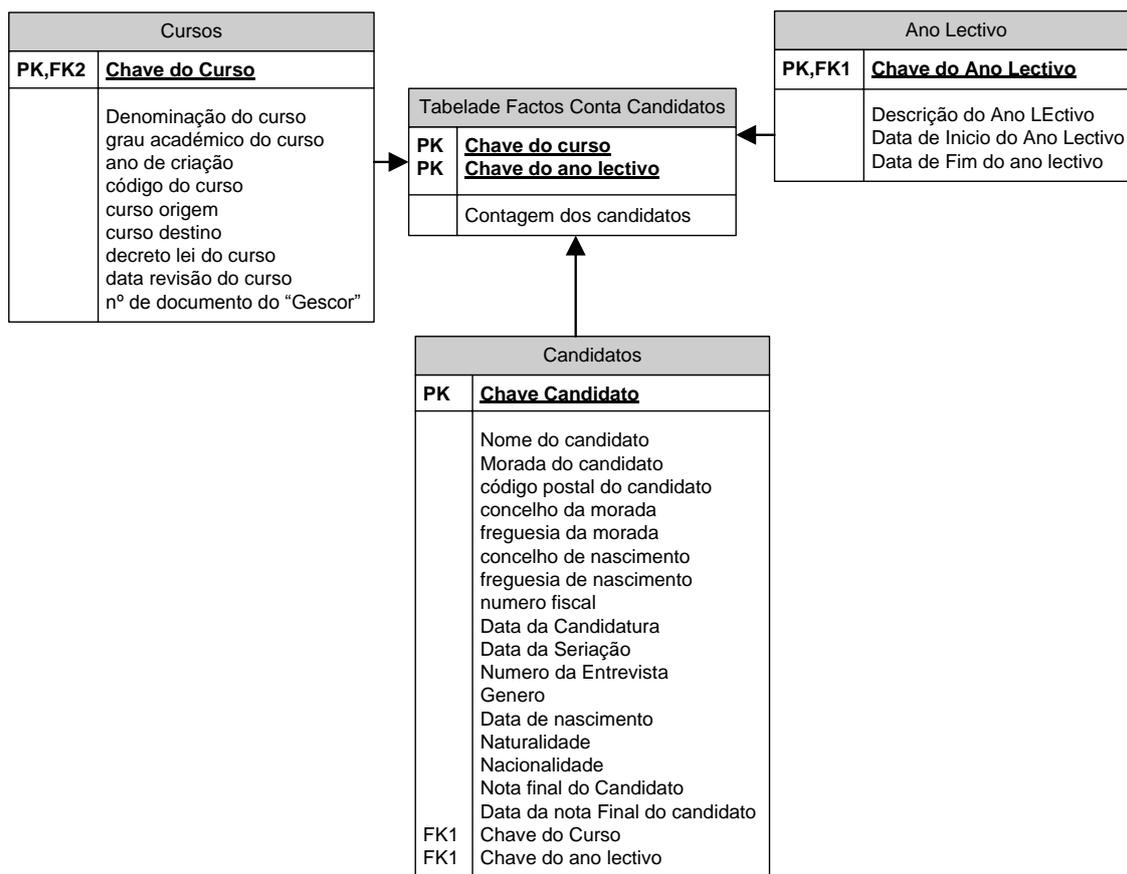


Figura 14 - Modelo Estrela - Fonte: Elaboração própria

A modelação multidimensional permite apresentar os dados de forma estruturada e intuitiva para que o seu acesso seja feito com grande performance. Este modelo, segundo Kimball, contém três conceitos básicos: factos, dimensões e métricas (Kimball, L., M., & W., 1998).

2.3.3. FACTOS

Um facto representa as informações que serão analisadas, sendo formado normalmente por valores numéricos que representam dados de medida. É uma colecção de itens de dados relacionados, uma medida ou uma leitura de um acontecimento, originado por um determinado processo. Regra geral, cada facto representa um item ou transacção de negócio. Os factos são inseridos naquilo que designamos como tabela de factos e podem apresentar-se de forma simples ou agregada.

A tabela de factos é considerada a principal tabela do modelo dimensional, onde estão e serão armazenadas as medições numéricas de interesse para a organização. Representam transacções, acontecimentos, ou outras actividades, que são utilizadas para medir o funcionamento e os resultados dos processos de negócio.

A tabela de factos é composta pelas Chaves Estrangeiras, que as ligam às dimensões, por uma chave primária (formada por uma combinação única de valores de chaves de dimensão) e pelas métricas de interesse para o negócio.

Outra característica para identificar um facto é sua natureza evolutiva, mudando as suas medidas no tempo podendo ser questionadas. Consideremos o seguinte facto como exemplo: A nota de um aluno, numa unidade curricular, num ano lectivo, num determinado momento de avaliação.

Tabela Facto	
PK	<u>Codigo Aluno</u>
PK	<u>Codigo Data</u>
PK	<u>Codigo Disciplina</u>
PK	<u>Codigo Momento Avaliação</u>
PK	<u>Codigo Ano Lectivo</u>
	Nota

Chaves Estrangeiras

Factos

Figura 15 - Exemplo de tabela de facto - Fonte: Elaboração própria

Segundo (Caldeira, Data Warehousing, Conceitos e Modelos., 2008, p. 63) “A tabela de factos é o elemento principal no modelo dimensional.”

Conforme (Kimball, L., M., e W., 1998), não se deve preencher uma linha da tabela facto com zeros para representar que nada aconteceu (por exemplo, se não houve vendas numa data especifica para um cliente, não se preenchem os dados), pois isso faria com que a tabela de factos crescesse demais.

Os factos podem ser considerados:

- Aditivos: Quando os valores são passíveis de serem somados em todas as dimensões;

- Semi-Aditivos: Quando a soma pode ser efectuada relativamente a algumas dimensões ou exclusivamente a uma dimensão;
- Não-Aditivos: Quando o valor não puder ser somado em nenhuma dimensão, ou produzir qualquer valor sem sentido válido.

As tabelas de factos subdividem-se em dois grandes grupos:

- Tabelas de factos propriamente ditas, são aquelas registam medidas, e que podem ainda subdividirem-se em tabelas de factos transaccionais, tabela de factos de sumarização periódica e finalmente as tabelas de factos de sumarização acumulada.
 - Tabelas de factos transaccionais, registam os factos que se vão medindo em múltiplos pontos discretos ao longo de um determinado período de tempo;
 - Tabela de factos de sumarização periódica, agregam as métricas de um processo de negócio em intervalos de tempo regulares e pré-estabelecidos;
 - Tabela de factos de sumarização acumulada, este tipo de tabelas é independente do tempo e cobre o ciclo de vida completo de um acontecimento (Caldeira, Data Warehousing, Conceitos e Modelos., 2008);
- As tabelas de factos sem factos caracterizam-se por não possuírem factos ou quando muito apenas servem para contar registos (Exemplo: Nº de presenças na disciplina). Apenas marcam a existência de um acontecimento.

2.3.4 DIMENSÕES

Uma dimensão é uma colecção de atributos textuais, que descrevem os objectos da organização ou as dimensões de negócio, e que se encontram altamente relacionados uns com os outros. As dimensões são homogéneas dado que armazenam apenas um único tipo de entidade (por ex. alunos, docentes, campus, cursos,...).

Cada tabela de dimensão tem uma chave primária que corresponde a um dos componentes da chave composta da tabela de factos. Como refere (Caldeira, Data

Warehousing, Conceitos e Modelos., 2008, p. 87), “*Como estas tabelas descrevem as entidades e objectos que participam nos processos de negócio de uma organização, o seu papel é vital em todo o ambiente do Data Warehouse.*”

Uma das principais funções dos atributos de tabelas de dimensão é servir como fonte para restrições numa consulta ou como cabeçalhos de linha utilizadas em queries e nos mapas impressos de resposta do utilizador.

As dimensões são os elementos que participam num facto, ou seja, as possíveis formas de visualizar os dados, por isso normalmente possuem poucos ou nenhuns atributos numéricos.

2.3.5. MÉTRICAS OU MEDIDAS

Uma medida é um atributo numérico de um facto o qual representa o desempenho ou o comportamento do negócio relativo as dimensões que participam desse facto. Esses números são denominados de variáveis. Como exemplos de medidas temos: Valor total de Propinas, número de alunos inscritos num curso, número de bolsas de estudo concedidas, número de diplomados, número de alunos desistentes.

A medida é determinada pela combinação das dimensões que participam num facto e estão localizadas como atributos na tabela de factos.

2.4. ESQUEMAS DO MODELOS DE DADOS DIMENSIONAL

Após a descrição anterior dos conceitos associados ao modelo de dados dimensional, de seguida irei apresentar os três tipos de esquemas:

- Esquema em estrela;
- Esquema em floco de neve;
- Esquema em constelações de factos.

2.4.1 ESQUEMA EM ESTRELA

O esquema em estrela está relacionado com a disposição das tabelas no modelo, onde no centro existe a tabela de factos, que se relaciona com as tabelas de dimensão que se encontram em seu redor.

O esquema em estrela é uma estrutura constituída por uma tabela de factos associada a várias tabelas de dimensão. A simetria do desenho e a simplicidade semântica são as qualidades mais notáveis neste tipo de modelo de dados (Caldeira, Data Warehousing, Conceitos e Modelos., 2008).

A sua composição típica mostra uma grande entidade central denominada facto, e um conjunto de entidades menores denominadas dimensões, distribuídas em redor dessa entidade central, formando uma estrela como mostra a figura 16.

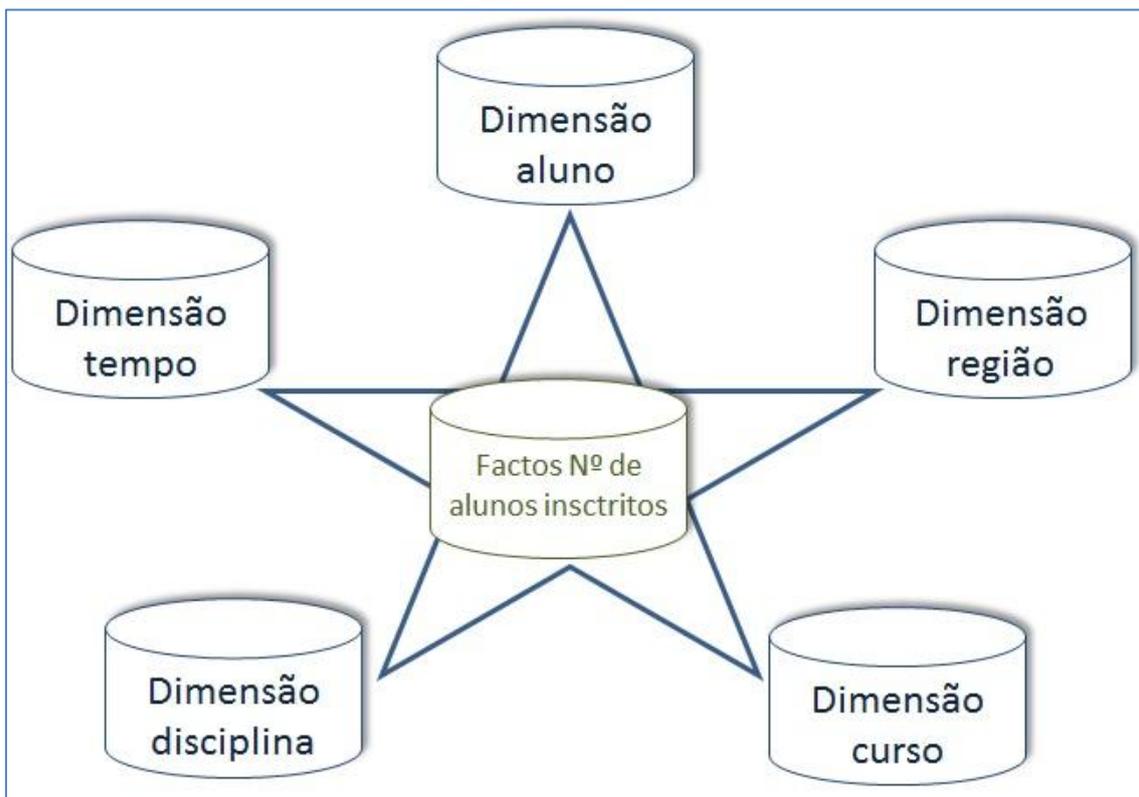


Figura 16 - Esquema estrela - Fonte: Adaptado de Caldeira (2008)

2.4.2 ESQUEMA EM FLOCOS DE NEVE

O esquema em flocos de neve, resulta de um refinamento do esquema em estrela com algumas dimensões normalizadas. onde cada uma das pontas da estrela passa a ser o centro de outras estrelas.

No fundo este modelo é o resultado da aplicação da terceira forma normal sobre as entidades dimensão.

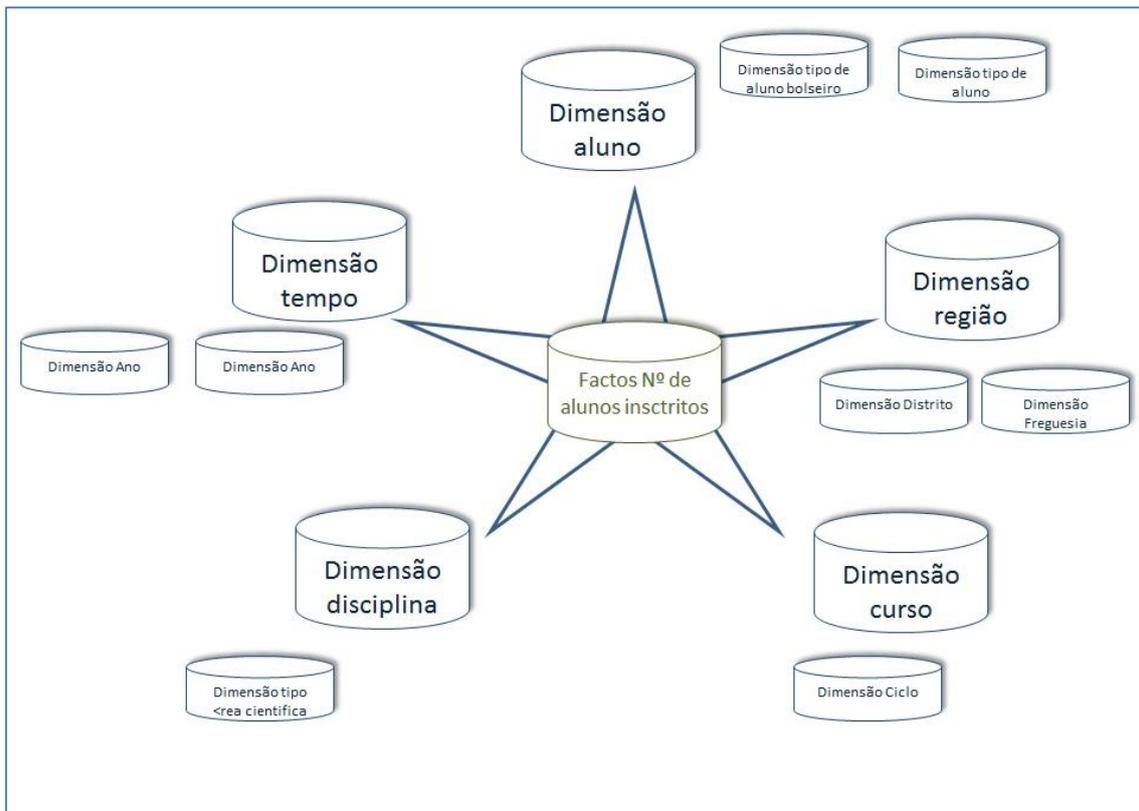


Figura 17 - Esquema em Floco de neve Fonte: Adaptado de Caldeira (2008)

2.4.3 ESQUEMA EM CONSTELAÇÃO

O esquema constelação possui múltiplas tabelas de factos e várias dimensões partilhadas por essas tabelas, constituindo um conjunto de estrelas.

De uma forma geral, os *Data Mart's* utilizam o modelo estrela e *Data Warehouse* utiliza o modelo constelação, que como pode ser visto na figura abaixo é a união de várias estruturas de estrela.

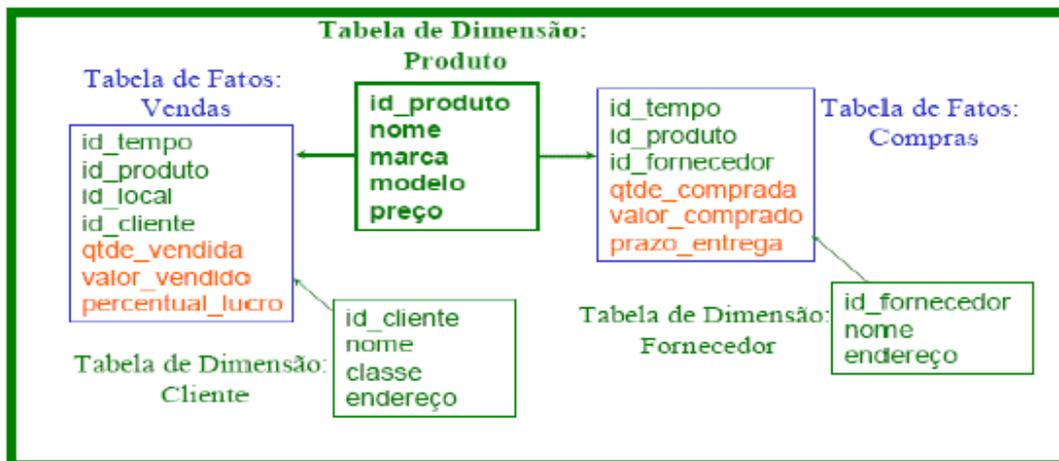


Figura 18 - Modelo em constelação- Fonte: Adaptado de Kimball (1996)

2.5. DATA MART

Quando se fala em *Data Warehouse*, surge um outro termo, o *Data Mart* (DM). Um *Data Mart* é em tudo idêntico a um DW, com a diferença que o *Data Mart* tem dados de uma área funcional e um *Data Warehouse* tem dados de toda a organização. Representa um sub-conjunto de dados do DW.

Data Mart's (DM) são bases de dados departamentalizadas (unidades de negócio), que podem apresentar visões relacionais ou multidimensionais, é também utilizado por alguns autores para se referir a um DW de pequena capacidade, usado para responder especificamente a uma unidade de negócio ou a um departamento da organização.

Similarmente ao DW, o *Data Mart* pode conter dados armazenados em vários níveis de detalhe, dependendo das funções do utilizador final e das exigências do negócio.

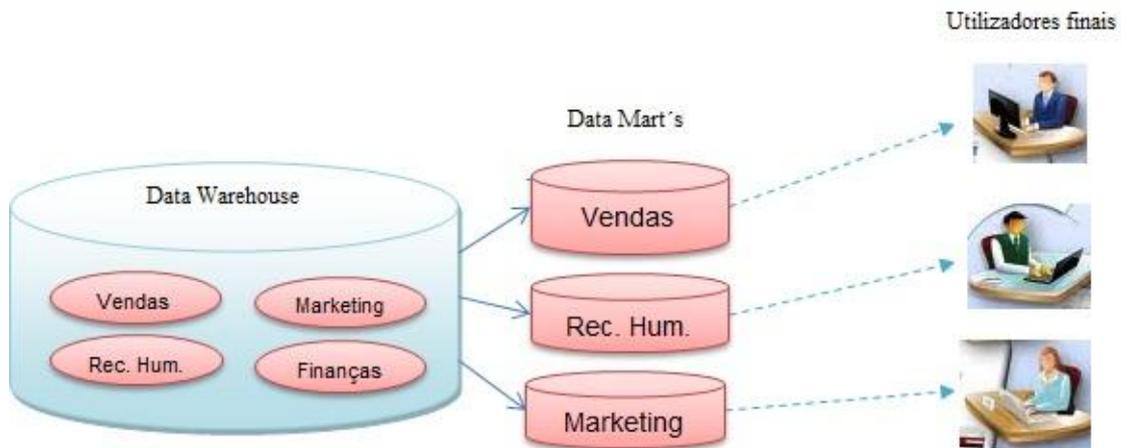


Figura 19 - Exemplo de Data Mart - Fonte: Elaboração própria

Resumindo, o termo *Data Mart* representa, um depósito de dados direccionado com um sector ou departamento específico da organização. Podemos dizer que um DM é um DW sectorial. Segue as mesmas características de um DW geral, mas os seus dados são orientados apenas para um sector. Desta forma podemos também considerar que um DW é formado por vários DM's.

Os *Data Marts* podem ser dependentes quando detém os dados, relativos a um assunto, a partir de um DW ou, independentes quando os dados vêm directamente das fontes de dados internas ou externas à organização, sem passar pelo DW.

Há dois tipos básicos de *Data Marts*:

- Dependentes do *Data Warehouse*, quando o *Data Mart* actualiza seus dados através do DW;
- Independentes do *Data Warehouse*, quando o *Data Mart* actualiza os seus dados através dos dados operacionais.

3. METODOLOGIA DO PROJECTO DW

Inmon e Kimball são referências nesta área, dado que cada um é criador de uma escola de pensamento e de prática de DW. Neste capítulo apresentamos de forma sucinta as diversas fases das suas metodologias.

Na criação do DW e dos DM's, Inmon propõe uma metodologia de desenvolvimento do tipo *top-down*, onde sugere que o DW deva ser projectado de forma una, modelando-se a organização como um todo, chegando-se desta forma a um único modelo, passando a partir daqui à construção dos *Data Mart's*.

Os dados são transformados e organizados temporalmente numa única base de dados (*Data Warehouse*). De seguida, são criadas, através de processos de extracção, BDs menores com parte dos dados do DW, denominadas *Data Mart's* (DM), onde os utilizadores finais realizariam análises ad hoc, relatórios automáticos e *dashboards*.

Ralph Kimball recomenda uma metodologia inversa à de Inmon, que preconiza uma arquitectura de múltiplas bases de dados designadas por DM's que estão organizadas por processos de negócio, onde a soma destas bases de dados e sistemas individuais dá origem ao DW. Desta forma este autor recomenda uma metodologia de desenvolvimento do tipo bottom-up.

3.1 KIMBALL

O método Kimball é independente tanto do fornecedor de software como do de hardware, mas surge da prática e da experiência do seu autor em projectos de desenvolvimento de sistemas de *Data Warehouse*. Os DWs construídos com o modelo Kimball usam a Modelação Dimensional (MD).

Neste método, outra característica importante é a arquitectura global ser caracterizada por múltiplas BDs, designadas por *Data Marts*, que são interligadas entre si pelo *Data Warehouse Bus*, responsável pela coerência da informação integrante do DW.

Este autor defende o ciclo de desenvolvimento do DW em 12 fases conforme apresentado na Figura 20 retirada de (Kimball, 1996).

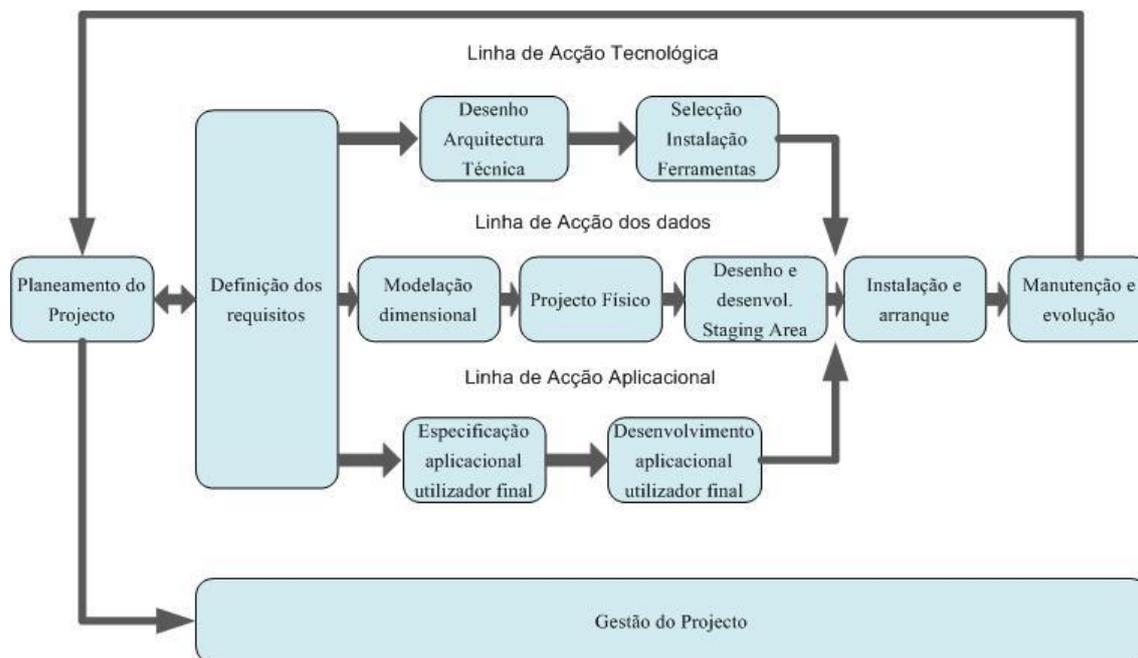


Figura 20 - Modelo de implementação Fonte: Adaptado de Kimball (1996)

Planeamento do Projecto

Nesta fase, correspondente à etapa inicial de desenvolvimento do DW, pretende-se obter conhecimento sobre a organização: os processos; as prioridades de negócio; a identificação das necessidades e a identificação da equipa de projecto. Ainda nesta fase, a elaboração do plano geral do projecto deve estabelecer um prazo para o mesmo, deve mencionar os recursos necessários e as expectativas dos utilizadores. Como resultado desta actividade temos o desenvolvimento de uma versão preliminar do âmbito do projecto.

Definição dos requisitos

Nesta fase, pretende-se identificar e desenvolver os requisitos iniciais do Sistema de *Data Warehouse*, pretende-se efectuar o apuramento e identificação dos requisitos necessários. Esta fase pode ser feita de forma individual concentrada no utilizador/grupos de utilizadores ou como segunda alternativa realçando os dados existentes nos sistemas aplicativos da organização.

O levantamento de requisitos guiado pelo utilizador é, normalmente, realizado através do levantamento das funções executadas pelos utilizadores nas entrevistas e reuniões.

O método da obtenção dos requisitos, a partir dos dados existentes na organização, é realizado através da análise dos modelos de dados existentes nas BD's ou dos dados existentes de forma avulsa.

Esta especificação inicial, designada por especificação dos requisitos, irá servir como ponto de partida de três grupos de actividades que decorrerão em paralelo: Grupo A (A1-Definição da arquitectura Tecnológica, A2- Selecção e Instalação de Ferramentas); Grupo B (B1- Modelação Dimensional, B2- Projecto Físico, B3- Desenho e Desenvolvimento área de Estágio) e Grupo C (C1-Especificação Aplicacional Utilizador Final, C2-Desenvolvimento Aplicacional Utilizador Final).

Definição da Arquitectura Tecnológica

São definidos critérios para a selecção dos produtos (Equipamentos, Hardware, Sistemas de Gestão de Bases de Dados, Sistemas de carregamento de dados).

Seleção, Instalação de Ferramentas

Actividade responsável pela selecção e instalação do definido em Arquitectura Tecnológica.

Modelação Dimensional

A definição dos requisitos determina os dados necessários às exigências analíticas do negócio: a identificação e escolha dos factos, a sua granularidade, as medidas a partir do conjunto de requisitos e a escolha das dimensões.

Projecto físico

Os requisitos para esta etapa incluem a base de dados, a desnormalização de algumas estruturas de dados, a padronização de uso de nomes para os objectos da BD e acções que visem melhorar o desempenho e performance.

Desenho e desenvolvimento da área de estágio

Esta etapa descreve o desenho e construção da *staging área*, a extração, transformação e carga dos dados a partir de uma ou mais fontes de dados.

Deverão também aqui ser detalhados quais os dados a serem extraídos, transformados e carregados, como serão mantidos, quais as ferramentas ou mecanismos que serão utilizados para automatizar este processo, detalhar e/ou esquematizar como serão povoadas as tabelas dimensão e factos, a sua atualização e manutenção, prevendo meios eficientes de carga.

Especificação e desenvolvimento Aplicacional – Utilizador Final

Estas duas etapas descrevem as aplicações que virão a ser utilizadas para aceder ao sistema de *Data Warehouse*.

Instalação e Arranque

Antes da instalação deve ser feito um *check-list* da funcionalidade da plataforma, seguidos de testes e do agendamento da instalação final do sistema. Nesta actividade, há que garantir que o sistema esteja construído e carregado com os registos informacionais. É nesta actividade que se define a estratégia de formação a dar aos utilizadores.

Manutenção e evolução

Nesta fase de manutenção, acompanhamento e evolução devem ser incluídas técnicas para avaliar o ambiente do DW, verificar: se as necessidades estão a ser atendidas, se os dados estão a ser actualizados numa correcta escala de tempo que evolução se pretende para o DW, entre outros aspectos.

Gestão do Projecto

Nesta actividade deve ser criada uma comissão que monitoriza, gere e define as prioridades do projecto e que garante que todas as actividades que apoiam o ciclo de vida do projecto estão devidamente sincronizadas e “em linha”. Esta

actividade deverá incluir o desenvolvimento dum plano de comunicação entre os departamentos/áreas de negócio envolvidas e o departamento de sistemas de informação.

3.2 INMON

A arquitectura proposta por Inmon contém todos os SIs e BDs da organização. O conjunto, designado por *Corporate Information Factory (CIF)*, divide o ambiente de BDs global da organização em quatro níveis: Operacional, *Data Warehouse*, Departamental e Individual.

O primeiro nível, designado por Operacional, como o próprio nome indica contém os dados dos sistemas transaccionais e regista a actividade quotidiana da organização. É destes sistemas que os dados são extraídos, transformados e carregados no segundo nível “*Data Warehouse*”. O nível Departamental corresponde a resumos do DW e dependendo das exigências de informação, assim será a agregação dos dados. O nível Individual corresponde aos utilizadores individuais e acontece quando estes separam conjuntos *ad hoc* de dados como parte da análise necessária ao suporte das tomadas de decisão.

A arquitectura de Inmon assegura que todos os dados são consistentes porque todos os dados departamentais provêm do mesmo DW.

Para Inmon, só quando os três níveis da modelação existentes no sistema de análise top-down estão completos é que o desenvolvimento do DW pode ser iniciado.

O primeiro nível designa-se por Diagrama Entidade-Relação (DER) ou *Entity Relationship Diagram (ERD)*. A equipa de desenvolvimento cria um DER para cada departamento que vai usar o DW, à semelhança do que faria para uma BD transaccional. O DER do DW é o resultado da integração de todos os DER's departamentais contidos no âmbito do projecto.

O segundo nível de modelação estabelece o modelo *Data Item Set (DIS)* para cada entidade do modelo DER do departamento. A integração dos vários DIS departamentais compõe o DIS global do DW. O modelo de dados deste nível tem quatro construtores:

- Agrupamento primário de dados – Apenas existe uma vez para cada entidade do modelo DER do *Data Warehouse*. Possui atributos que só existem uma vez para cada entidade do modelo DER e chaves que o relacionam com outros agrupamentos e tipos de dados.
- Agrupamento secundário de dados – Contém atributos que podem existir várias vezes para cada entidade do modelo DER. Devem existir tantos agrupamentos secundários como grupos de dados que se repetem.
- Conector – Representa a relação entre entidades do modelo DER.
- Tipo de dados – Contém os tipos de dados que podem ser gerados na entidade do modelo DER.

O nível mais baixo do modelo de dados é o físico. Segundo (Inmon W. , 2002):

"...The physical data model is created from the midlevel data model merely by extending the midlevel data model to include keys and physical characteristics of the model. "

Completar os três níveis de modelação é condição prévia para usar a metodologia de Inmon, mas esta metodologia não é mais que uma adaptação do desenvolvimento em espiral de Boehm (1988), que denominou Meth2, sendo Meth1 para desenvolver Sistemas Transaccionais e Meth3 para optimização de DWs existentes). Inmon, designou, por Decision Support 1 (DSS1), o primeiro passo da modelação de três níveis. A metodologia é constituída por dez passos, representados na Figura 21 e Tabela 5.

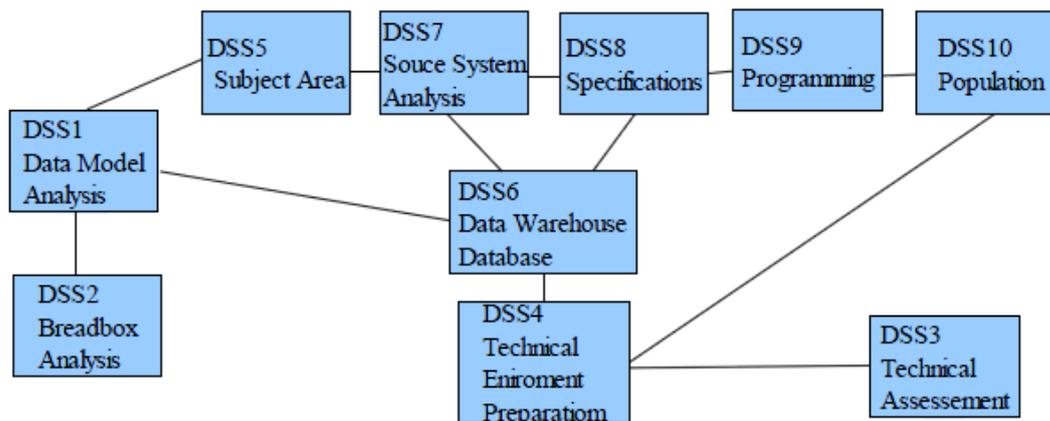


Figura 21 - Etapas da Metodologia - Fonte: Adaptado de Inmon (1997)

Usando o modelo de três níveis de dados como ponto de partida, o próximo passo é a análise de granularidade (DSS2) designada por *Breadbox Analysis*. A granularidade é a medida do detalhe dos dados. Os dados transaccionais têm o nível de granularidade mais baixo porque têm a informação mais detalhada possível. Este nível tem também a designação de atómico. Se o volume de dados for muito grande, então a equipa de desenvolvimento deve considerar diferentes níveis de granularidade para os dados (Inmon, 2002). Isto envolve armazenar alguns dados ao nível da transacção e outros em formas resumidas (ex: total diário, vendas anuais).

Quando a questão da granularidade estiver resolvida, é seleccionada a primeira área de negócio (DSS5). Esta vai tornar-se a primeira BD departamental a ser criada. A equipa de desenvolvimento analisa os sistemas transaccionais que fornecem os dados da primeira área de negócio (DSS7), define as especificações (DSS8), programa as especificações (DSS9) e carrega a BD (DSS10). Entretanto, o DW global começa a ser desenvolvido paralelamente (DSS6) e quando houver informação suficiente é efectuada uma avaliação técnica (DSS3).

Com o desenvolvimento sucessivo das BD departamentais, o DW é afectado pelas alterações. Na Figura 21 podemos ver o aspecto repetitivo das alterações representado

pelos linhas que ligam os vários passos ao DW. As linhas ligam a análise dos sistemas operacionais (DSS7) e as especificações (DSS8) com o DW (DSS6). Isto significa que o DW será revisitado cada vez que uma BD departamental nova é desenvolvida. A linha que liga o carregamento da BD departamental (DSS10) com a preparação do ambiente técnico (DSS4) demonstra a natureza repetitiva de Meth2. Com a preparação do ambiente técnico (DSS4), Inmon refere-se à verificação dos recursos de rede, isto é hardware de armazenamento, o sistema operativo e o software de acesso devem estar prontos para receber dados (Inmon, 2002).

A metodologia de Inmon pode ser resumida de acordo com a Tabela 5.

Etapa	Descrição
1- Análise do modelo de dados	Elaborar o modelo de dados para o DW.
2- Dimensionamento	Estimar o volume de dados que o DW vai conter.
3- Avaliação técnica	Definir as configurações técnicas para o DW
4- Preparação do ambiente técnico	Identificar tecnicamente como a configuração definida será acomodada.
5- Análise das áreas de interesse	Selecionar a área de negócio a ser povoada.
6- Projecto do <i>Data Warehouse</i>	Elaborar o projecto físico de base de dados para o DW.
7- Análise do sistema fonte	Identificar, nos sistemas existentes, a fonte de dados para o DW e efectuar um mapeamento entre eles.
8- Especificações	Definir, em especificações de programas, as rotinas para o carregamento dos dados.
9- Programação	Codificar as especificações para as rotinas de carga.
10-Povoamento	Executar as rotinas de carga do DW.

Tabela 5 - Metodologia para desenvolvimento de um DW (Inmon).

Na metodologia de Inmon, a fase 1 do projecto inicia-se com a Análise do modelo de dados que contempla um plano para a migração de dados.

Na fase 2 de Dimensionamento e a partir do modelo de dados estima-se qual o volume de informação planeada para o DW.

Em paralelo deve ser preparada a Avaliação Técnica (Fase 3) e a Preparação do Ambiente Técnico. A fase 3 contém a definição das configurações técnicas para o DW e arquitectura tecnológica de suporte de acordo com os requisitos estipulados para o DW. A fase 4 inclui a rede informática, o sistema de gestão para o DW, a construção do próprio DW, a definição e o interface com o DW. Esta tarefa é executada apenas uma vez.

Na Fase 5, Análise das áreas de Interesse, deve ser efectuada a escolha do tema para ser implementado, o qual deve ser abrangente para que a sua implementação tenha algum impacto na organização, mas não deve ser demasiado abrangente que torne o seu desenvolvimento uma tarefa impossível.

Na fase 6, Projecto do *Data Warehouse*, que é executada apenas uma vez na elaboração do desenho da BD e que pretende elaborar o projecto físico de base de dados para o DW.

Na fase 7, Análise do sistema fonte, é necessário analisar os sistemas fonte existentes, identificar qual a origem dos dados a transferir para o DW e efectuar um mapeamento entre eles.

Na fase 8 Especificações, é necessário definir, especificações de programas, as rotinas para o carregamento dos dados, definir as ferramentas que irão ser responsáveis pelas operações de extracção, transformação e integração dos dados provenientes das diversas fontes de informação da fase 7.

Na fase 9, Programação, são desenvolvidos os programas para extrair e carregar os dados escolhidos durante a fase 8.

Na fase 10 Povoamento, são executadas as rotinas de carga do DW.

3.3 INMON VERSUS KIMBALL

Inmon e Kimball criaram um grande debate em torno desta temática e apresentam duas abordagens diferentes no projecto de criação de um DW. A Tabela 6, apresenta as principais vantagens e desvantagens das duas metodologias:

Vantagens

Bill Inmon	Ralph Kimball
Melhor definição estratégica do projecto.	Infra-Estrutura mais adequada às exigências de um SAD.
Proporciona um recenseamento integral dos sistemas fonte e conteúdos existentes na organização.	Abordagem Iterativa centrada nas necessidades de informação. Permite antecipar a entrega de resultados.
Desenvolve uma abordagem sistematizada e completa sobre os processos de integração.	Garante o maior envolvimento dos utilizadores.
Metodologia mais vocacionada para definição das componentes de <i>back-end</i> .	Permite fasear os custos de investimento em infra-estrutura.
	Abordagem de implementação totalmente integrada.

Tabela 6 - Vantagens metodologias Inmon e Kimball

Desvantagens

Bill Inmon	Ralph Kimball
Abordagem <i>Top-Down</i> centrada nos dados, mais morosa e dispendiosa.	Dificuldade em definir as dimensões e factos conformes
Maiores custos iniciais em TI.	Vertiginoso crescimento do volume de dados armazenado.
Abordagem excessivamente centrada nos dados (todo o processo de desenvolvimento depende da prévia conclusão do modelo corporativo dos dados): - inviabiliza o envolvimento dos utilizadores no projecto; - prolonga o período de ausência de resultados. - relega para segundo plano a identificação das reais necessidades de informação dos utilizadores.	Conduz à obtenção de procedimentos de ETL, mais complexos: - modelos dimensionais requerem operações adicionais de transformação e agregação dos dados dos sistemas operacionais (usualmente representados em modelos normalizados); - alterações ao nível dos sistemas operacionais implicam alterações em procedimentos dedicados a diferentes esquemas em estrelas de diferentes granularidades.
Modelos normalizados pior desempenho analítico, sendo menos adequados e intuitivos para os SAD.	
Processo de desenvolvimento fragmentado: - primeiro o DW. - depois os <i>Data Marts</i>	

Tabela 7 - Desvantagens metodologias Inmon e Kimball

3.4 MICROSOFT

A *Microsoft* identifica um conjunto de componentes e actividades para o desenvolvimento deste tipo de projectos.

Assim, as componentes são:

1. Sistemas Fonte existentes – são os sistemas que armazenam todos os registos informacionais operacionais existentes de momento na organização. Nesta componente, são identificados e descritos os sistemas de onde vamos recolher a informação para o DW.
2. Sistema de Armazenamento – é o repositório físico final de dados. Existem no mercado um conjunto de sistemas que podem ser candidatos a receber o DW, contudo devem prever o armazenamento dos registos em ambiente multidimensional, nomeadamente os esquemas em estrela, com informação agregada em estruturas *OLAP*.
3. Ferramentas de exploração - são as ferramentas tecnológicas que permitem aos utilizadores explorar e aceder à informação no *Data Warehouse*. Estas ferramentas devem também ser identificadas por tipo de utilizadores, como por exemplo:
 - a1) os que só lêem relatórios;
 - a2) os que produzem relatórios;
 - a3) analistas de negócio ou super-utilizadores;
 - a4) gestores (executivos).
4. Ferramentas de modelação e desenvolvimento – consistem nas aplicações informáticas utilizadas para construir os ambientes de exploração.
5. Ferramentas de manutenção: são aplicações informáticas utilizadas para gerir o desempenho, segurança e integridade do Sistema de *Data Warehouse*.

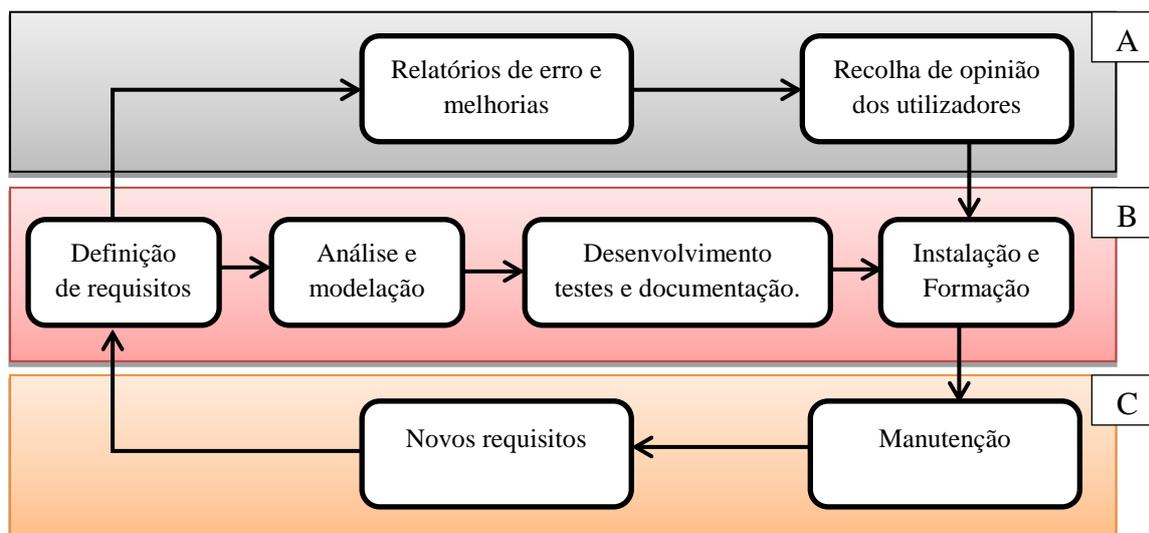


Figura 22 - Metodologia de implementação Microsoft- Fonte: Elaboração própria

O método inclui a execução de oito actividades onde quatro estão direccionadas para a implementação do Sistema de *Data Warehouse* (B), duas para as tarefas de revisão (A) e as restantes duas para as tarefas de manutenção (C). O processo é iterativo, isto é, o resultado das actividades de revisão e de manutenção fornecerão indicações ou recomendações para as próximas actividades de construção.

As quatro actividades para a implementação são:

1. Definição de requisitos – nesta actividade tenta-se identificar e perceber o negócio da organização e ainda o sector onde está inserida. Os requisitos são recolhidos a partir de entrevistas ou conversas informais a diversos tipos de utilizadores, por exemplo: directores de unidades, gestores e utilizadores dos sistemas de informação;
2. Análise e modelação – os registos operacionais são identificados e documentados, as áreas de armazenamento são modeladas, o processo de extracção dos registos informacionais operacionais para o *Data Warehouse* são desenvolvidos, e é definido o ambiente de exploração;
3. Desenvolvimento, testes e documentação – o repositório do *Data Warehouse* é construído e carregado com os registos informacionais operacionais;

4. Instalação e formação – o Sistema de *Data Warehouse* é disponibilizado e é efectuada a formação dos utilizadores.

As restantes quatro actividades podem ser divididas em dois grupos o de manutenção e o de revisão:

1. Manutenção e funcionamento do *Data Warehouse* – o objectivo é monitorizar o funcionamento do *Data Warehouse* garantindo o seu funcionamento e manutenção e que o sistema está a corresponder às exigências para que foi construído.

2. Revisão sobre o Sistema de *Data Warehouse* – pretende identificar a percepção e a reacção que os utilizadores têm da utilização do sistema.

Em qualquer uma das fases, anteriormente descritas, podem surgir novos requisitos operacionais, que deverão ser identificados, documentados e utilizados como base para futuros desenvolvimentos, alimentado assim futuras actividades de definição de requisitos.

4. ORGANIZAÇÃO DO ENSINO SUPERIOR E O IPT

O objectivo deste capítulo é descrever a organização do ensino superior em Portugal, e em particular o Instituto Politécnico de Tomar, organização onde se desenvolve este projecto, enquadrando as actividades de dois organismos de avaliação: uma agência internacional, a EUA e outra nacional a A3E's Agência de Avaliação e Acreditação do Ensino Superior.

4.1 INTRODUÇÃO

A universidade é considerada uma instituição secular e possui como missão assegurar uma sólida preparação científica e cultural que fomente as capacidades de concepção, de inovação e de análise crítica. É um lugar, de estudo, investigação e reflexão e ainda deve possuir as condições para o surgimento de grandes propostas para a compreensão e transformação do mundo. As universidades também possuem como missão a gestão do mais precioso património de que a humanidade dispõe, o dos saberes acumulados durante séculos de observação e de pensamento.

Por outro lado, devido à sua própria designação, o politécnico visa proporcionar uma sólida formação cultural e técnica de nível superior, desenvolver a capacidade de inovação e de análise crítica e ministrar conhecimentos científicos de índole teórica e prática e as suas aplicações, com vista ao exercício de actividades profissionais. Será o lugar para se formarem técnicos, altamente especializados, possuidores de um treino que lhes possibilite a criação de técnicas inovadoras, sendo indispensável a investigação, desenvolvida tanto nas universidades como nos politécnicos.

Retém-se a ideia de uma lógica, preferencialmente profissionalizante, direccionada para o exercício de actividades profissionais do ensino politécnico e o desenvolvimento da capacidade de concepção indicada para o ensino universitário.

4.2 ORGANIZAÇÃO DO ENSINO SUPERIOR

O sistema de ensino português está organizado em quatro níveis de ensino:

1. Pré-escolar (de cariz não obrigatório, destina-se a crianças dos 2 aos 6 anos);

2. Básico (escolaridade mínima obrigatória, organiza-se em 3 ciclos de estudos, distribuídos por 9 anos);
3. Secundário (composto por 3 anos de ensino facultativo);
4. Superior (organizado em 3 ciclos de estudos – licenciatura, mestrado e doutoramento – habitualmente com 3, 2 e 3 anos de duração).

Em Portugal, o ensino superior tem como objectivo a qualificação de alto nível dos portugueses, a produção e difusão do conhecimento, bem como a formação cultural, artística, tecnológica e científica dos seus estudantes num quadro de referência internacional.

Actualmente, o sistema de ensino superior em Portugal apresenta-se sob a forma de uma estrutura dual, em que coexistem o ensino universitário e o ensino politécnico.

O enquadramento do ensino politécnico, no sistema de ensino superior ganhou a forma actual com a publicação da Lei nº 46/86, de 14 de Outubro, Lei de Bases do Sistema Educativo, onde se distingue o ensino universitário e ensino politécnico da seguinte forma:

- *“O ensino universitário visa assegurar uma sólida preparação científica e cultural e proporcionar uma preparação técnica que habilite para o exercício actividades profissionais e culturais e fomenta o desenvolvimento das capacidades de concepção, de inovação e de análise crítica”.*

- *“O ensino politécnico visa proporcionar uma sólida formação cultural e técnica de nível superior, desenvolver a capacidade de inovação e de análise crítica e ministrar conhecimentos científicos de índole teórica e prática e as suas aplicações com vista ao exercício de actividades profissionais”.*

Retém-se assim a ideia de uma lógica preferencialmente profissionalizante, direccionada para o exercício de actividades profissionais do ensino politécnico e o desenvolvimento da capacidade de concepção indicada para o ensino universitário (Chaves, 2010).

De um modo geral, as pessoas dão continuidade aos seus estudos, por exemplo, no ensino politécnico, para adquirirem competências técnicas, visando aceder a um estatuto profissional mais valorizado (Reis & Reis, 2008).

Por outro lado, a universidade é o local onde o discente adquire uma base cultural forte e aprende a reflectir autonomamente, sendo, assim, um espaço onde se exercita a procura do saber e se estimula a capacidade intelectual para pensar e resolver problemas (Reis & Reis, 2008).

Os cursos colocados à disposição dos interessados, pelas instituições e as respectivas unidades orgânicas conferem os seguintes graus:

Grau de Licenciado

As instituições que conferem este grau são as universitárias e politécnicas.

O ciclo de estudos conducente ao grau de licenciado, no ensino politécnico, tem uma duração normal de seis semestres curriculares de trabalho dos alunos correspondentes a 180 créditos. Excepcionalmente e em casos cobertos por normas jurídicas nacionais ou da União Europeia, os ciclos de estudo podem ter uma duração normal até sete ou oito semestres curriculares de trabalho e uma formação até 240 créditos.

O ciclo de estudos conducente ao grau de licenciado, no ensino universitário, tem 180 ou 240 créditos e uma duração normal compreendida entre seis e oito semestres curriculares de trabalho dos alunos. No 1.º ciclo de estudos das instituições universitárias ou politécnicas o grau de licenciado é conferido, através da aprovação em todas as unidades curriculares que integram o plano de estudos do curso de licenciatura, e tenham obtido o número de créditos fixado.

Grau de Mestre

As instituições universitárias e politécnicas conferem o grau de mestre.

O ciclo de estudos conducente ao grau de mestre tem 90 a 120 créditos e uma duração normal compreendida entre três e quatro semestres curriculares de trabalho dos alunos.

No ensino politécnico, o ciclo de estudos conducente ao grau de mestre deve assegurar, predominantemente, a aquisição de uma especialização de natureza profissional. No ensino universitário, o ciclo de estudos conducente ao grau de mestre deve assegurar, predominantemente, a aquisição de uma especialização de natureza académica com recurso à actividade de investigação ou que aprofunde competências profissionais.

No ensino universitário, o grau de mestre pode igualmente ser conferido após um ciclo de estudos integrado, com 300 a 360 créditos e uma duração normal compreendida entre 10 e 12 semestres curriculares de trabalho nos casos em que a duração para o acesso ao exercício de uma determinada actividade profissional seja fixada por normas legais da União Europeia ou que resulte de uma prática estável e consolidada na União Europeia. Neste ciclo de estudos, é conferido o grau de licenciado aos que tenham realizado os 180 créditos correspondentes aos primeiros seis semestres curriculares de trabalho.

No 2.º ciclo de estudos das instituições universitárias ou politécnicas, o grau de mestre é conferido aos que obtenham aprovação em todas as unidades curriculares que integram o plano de estudos do curso de mestrado e da aprovação no acto público de defesa da dissertação, referente ao trabalho de projecto ou relatório de estágio.

Grau de Doutor

O grau de doutor é conferido, pelas instituições universitárias aos que tenham obtido aprovação nas unidades curriculares do curso de doutoramento, quando exista e aprovação no acto público de defesa da tese.

Diagrama do Sistema de Ensino Superior Português

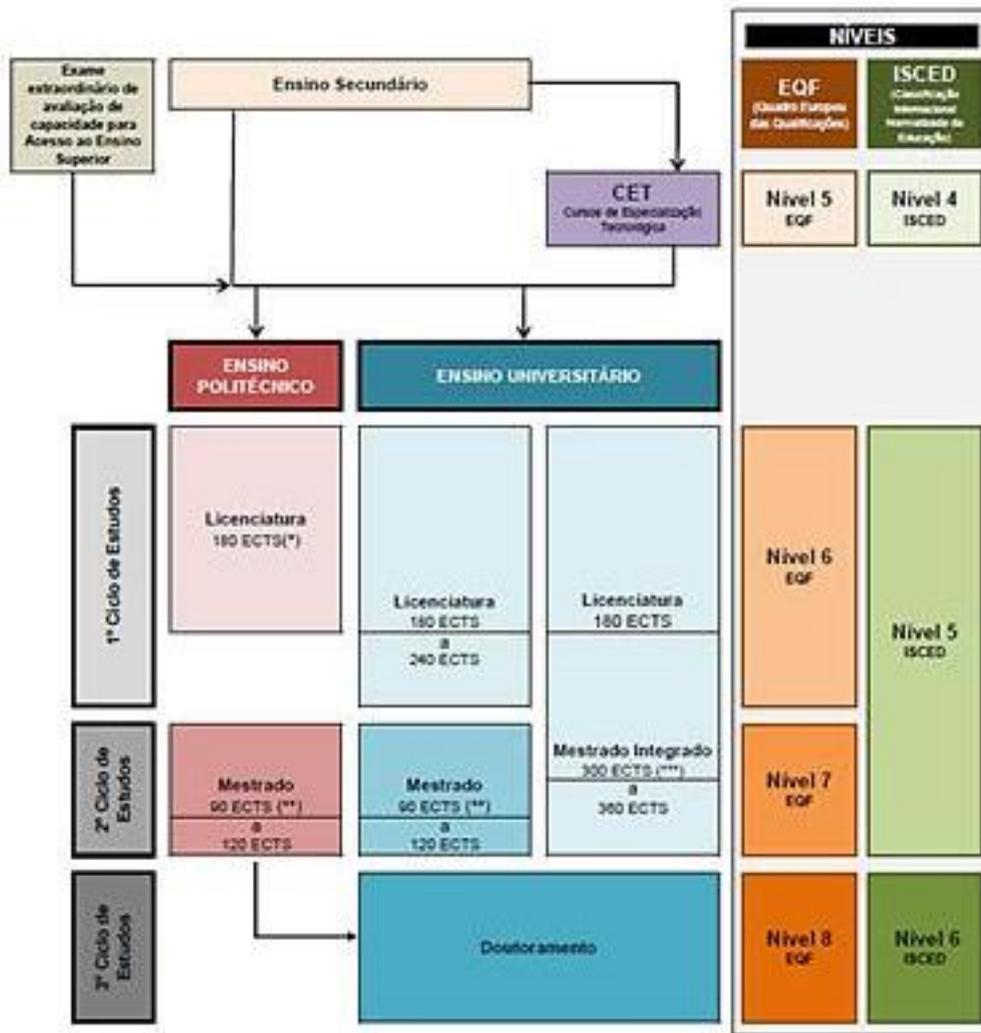


Figura 23 - Diagrama do Sistema de Ensino Superior Português - Fonte: DGES/MCTES (2011)

4.3 INSTITUTO POLITÉCNICO DE TOMAR

O IPT foi criado em 1 de Janeiro de 1997, mas, efectivamente, as actividades deste Instituto remontam a Outubro de 1982, de acordo com a seguinte evolução:

1. O Decreto-Lei 402/73, de 11 de Agosto, criou o IPT para o qual não foi nomeada uma Comissão Instaladora, pelo que não chegou a entrar em funcionamento; em Abril de 1979, foi criada a Escola Superior de Tecnologia de Tomar, Escola não integrada, cuja Comissão Instaladora tomou posse em 26 de Outubro de 1982;

2. O Decreto-Lei 46/85, de 22 de Novembro, integrou a Escola Superior de Tecnologia de Tomar no Instituto Politécnico de Santarém (IPS);
3. O Decreto-Lei 304/94, de 19 de Dezembro, criou a Escola Superior de Tecnologia e Gestão de Tomar (ESTGT), integrada no IPS, que sucedeu, para todos os efeitos legais, à extinta Escola Superior de Tecnologia de Tomar;
4. O carácter pluridisciplinar da ESTGT conferia-lhe vocação para adquirir o estatuto de Instituto Politécnico, o que veio a concretizar-se pelo Decreto-Lei 96/96, de 17 de Julho, que decretou a separação imediata desta Escola do IPS e a criação do IPT, que entrou em funcionamento no início do ano seguinte.

4.3.1 ORGANIZAÇÃO INTERNA

O IPT integra Escolas, unidades de formação não graduada (centros de estudos), serviços de acção social e serviços técnicos especializados, coordenados, orientados pelo Presidente e demais órgãos de governação e gestão do IPT consagrados pelos respectivos Estatutos.

4.3.1.1 Escolas

As Escolas do IPT são estruturas permanentes de recursos humanos e materiais dotadas de objectivos próprios. Cada Escola é organizada por Departamentos e Áreas Interdepartamentais que, no seu conjunto, ministram um ou mais cursos num domínio bem definido da actividade cultural ou profissional. São elas:

1. A **Escola Superior de Gestão de Tomar**, situada, em Tomar no Campus do IPT, foi criada em 1996, tal como o próprio IPT, mas a sua actividade remonta a 1986. Actualmente, dispõe de uma oferta formativa alargada na área da Gestão, aplicada aos seguintes ramos: Empresas, Turismo, Recursos Humanos, Saúde, Banca, Comércio, Serviços, Auditoria, Fiscalidade e Administração Pública.
2. A **Escola Superior de Tecnologia de Tomar**, situada no Campus do IPT, oferece formação nas áreas de Artes e Engenharia. Nas Artes, destacam-se os

ramos de Conservação e Restauro, Arqueologia, Artes Plásticas, Fotografia e Artes Gráficas. Nas Engenharias oferta formativa em Química e Bioquímica, Ambiente, Civil, Electrotécnica e Informática.

3. A **Escola Superior de Tecnologia de Abrantes**, situada no centro da cidade de Abrantes, oferece formação nas áreas de Comunicação, Cinema Documental, Tecnologias de Informação, Design e Engenharia Mecânica. Esta Escola foi criada e entrou em funcionamento em 1999.

4.3.1.2 Unidades de formação

As unidades de formação do IPT, garantem, em áreas de actuação específicas, o desenvolvimento de projectos nos domínios da formação, investigação e de prestação de serviços ao exterior, bem como outras actividades que se enquadrem na missão do IPT, através dos recursos humanos e materiais que lhe são atribuídos e em articulação com as Escolas.

Até ao momento, foram criados em parceria com as respectivas autarquias e encontram-se em funcionamento: o Centro de Estudos Politécnicos de Torres Novas (CEPTON), o Centro de Estudos Politécnicos da Golegã (CESPOGA), o Centro de Estudos Politécnicos da Sertã (CEPSES), o Centro de Estudos e Formação Especializada de Ferreira do Zêzere (CEFE.fz), o Centro de Formação Politécnica de Mação (CEFOPOM) e Instituto Terra e Memória (ITM), criado em parceria com a Câmara de Mação.

4.3.1.3 Serviços de acção social

O IPT integra Serviços de Acção Social para desenvolvimento do sistema de acção social escolar, preconizado por lei e ainda para apoio de actividades culturais, desportivas e outras, que garantam equidade no acesso ao ensino superior, a prática de uma frequência bem sucedida e o projecto de formação global dos estudantes. Esta unidade do IPT é dotada de autonomia administrativa, financeira e patrimonial.

4.3.1.4 Serviços técnicos especializados

O IPT dispõe de serviços que asseguram a qualidade, o dinamismo e a inovação em vários sectores da Instituição, em particular, no que respeita à coordenação e funcionamento dos cursos. Além disso, garantem a permanente ligação do IPT à comunidade exterior através de prestação de serviços. São eles: o Centro de Línguas (cl.ipt); o Centro de Sondagens e Estudos Estatísticos; o Centro de Incubação de Ideias e Negócios (CIN); o Centro de Formação à Distância; o Centro de Documentação e Arquivo (CDA); o Centro de Pré-História; o Gabinete de Avaliação e Qualidade; o Centro de Impressões e Publicações; o Centro de Artes e Imagem; o Centro de Estudos de Fotografia da Golegã e a Oficina de Transferência de Tecnologia e de Conhecimento (OTIC).

4.3.2 OFERTA FORMATIVA

Desde a sua génese, o IPT definiu uma estratégia de oferta educativa, norteadada pela combinação de cursos profundamente enraizados na realidade económica regional e cursos inovadores em termos nacionais e com ambição internacional (Arte e Arqueologia; Tecnologia e Artes Gráficas). Esta opção estratégica nunca se alterou e é a base da coerência actual do Instituto.

O IPT, através das suas Escolas e unidades de formação, dispõe, actualmente, das seguintes ofertas: 23 cursos de licenciatura; 7 cursos de mestrado; 3 cursos de formação pós-graduada e 16 cursos de CET. O IPT oferece ainda formação através de cursos, que se inserem nos programas de acção no domínio da aprendizagem ao longo da vida. Estes cursos são organizados pelos Departamentos e Áreas Interdepartamentais das Escolas e pelos serviços técnicos especializados, como o Centro de Línguas e o Centro de Sondagens e Estudos Estatísticos.

Actualmente, 3 728 estudantes usufruem da oferta formativa de 1º ciclo, 2º ciclo e CET das 3 Escolas e das unidades de formação do IPT, assegurada por 251 docentes e 176 funcionários não docentes. A tabela 8 identifica a distribuição da comunidade académica do IPT.

Comunidade	ESGT	ESTA	ESTT	Serviços Centrais	SAS	Total
Estudantes	1239	755	1734	-	-	3728
Docentes	66	55	130	-	-	251
Não docentes	10	12	35	86	33	176

Tabela 8- Distribuição Comunidade Académica IPT

4.4 ESTRUTURA ORGANIZACIONAL DO IPT

4.4.1 VALORES E MISSÃO

A missão do IPT enquadra-se na missão definida pela Lei n.º 62/2007, de 10 de Setembro, que define como vocação das instituições de Ensino Superior Politécnico “... a criação, transmissão e difusão da cultura e do saber de natureza profissional, através da articulação do estudo, do ensino, da investigação orientada e do desenvolvimento experimental”, atribuindo-lhe o poder de conferir os “graus de licenciado e mestre, nos termos da lei”.

Assim, o IPT pretende contribuir para a formação de recursos e de capital humano, em quantidade e perfil adequados à sociedade global do conhecimento, e participar em processos de produção e desenvolvimento do conhecimento científico e tecnológico. Esta dupla missão, sendo indissociável e inalienável, é vista num contexto com características próprias que a condicionam e que modelam os desafios que actualmente se colocam ao sistema de ensino superior. Entre essas características destacam-se:

1. A emergência de um mercado europeu de trabalho, que exige a produção, a comparação e o reconhecimento de qualificações, que criem condições gerais de mobilidade e empregabilidade;
2. O crescente dinamismo dos sistemas económicos e sociais e do próprio conhecimento, o aumento da mobilidade entre fronteiras e mercados, assim como, a volatilidade dos perfis formativos em ambientes cada vez mais competitivos, que implicam uma versatilidade e complexidade no perfil da oferta formativa, bem como, a necessidade de captar novos públicos e proporcionar a formação ao longo da vida;

3. O desenvolvimento de sistemas globais de tecnologias de informação e comunicação, conducentes a novas formas de organização de métodos pedagógicos e de relacionamento entre professores, alunos e investigadores e ao estabelecimento de redes no âmbito do ensino e da investigação;
4. A crescente competitividade, exigindo padrões de qualidade cada vez mais elevados, certificados no plano nacional e internacional, bem como a adopção de estratégias conducentes à implementação das melhores práticas, tanto em matéria de conteúdos formativos, como de organização e gestão, que implicam a introdução de sistemas de avaliação permanente;
5. A necessidade de transparência na utilização dos recursos públicos, que exigem a permanência de critérios de eficiência e eficácia para permitir um controle estrito e rigoroso dos custos de funcionamento dos sistemas de ensino superior, não descurando os princípios de equidade e de igualdade de oportunidades.

Tendo em vista o cumprimento da sua missão, o IPT norteia as suas linhas de orientação estratégica com base nas premissas enunciadas. Procura, deste modo, produzir conhecimento útil e formar indivíduos capazes de o compreender e utilizar na criação de valor, desenvolvendo, assim, capacidades, competências e aptidões, que os preparem para o mercado de trabalho e para o exercício de uma cidadania activa numa sociedade democrática.

4.4.2 OBJECTIVOS

Mantendo-se fiel aos seus propósitos, o IPT preparou um programa de orientação estratégica para o período 2007/2013, com a definição de metas e respectivos planos de acção, de indicadores globais, sectoriais e regionais e ainda de mecanismos de aferição de qualidade.

A estratégia definida incidu em três áreas fundamentais:

1. Reformulação do modelo organizacional e reposicionamento no contexto da oferta formativa nuclear;
2. Abertura ao exterior;

3. Alargamento da área de influência e extensão da oferta formativa.

Com base nas áreas referidas, foram definidos os seguintes objectivos estratégicos:

1. Desenvolver uma plataforma de aglutinação estratégica, de descentralização operacional e de autonomia científica e pedagógica, que articule, de forma matricial e efectiva, competências e produtos;
2. Reorganizar a oferta formativa nuclear do IPT, afirmando-a e diferenciando-a ao nível regional e nacional;
3. Reforçar a integração do IPT na sua envolvente regional através da consolidação de ofertas existentes e do desenvolvimento de novas ofertas;
4. Aprofundar a integração do IPT na rede nacional de ensino superior e nas redes nacionais de Investigação e Desenvolvimento (I&D);
5. Expandir e aprofundar a actuação do IPT na Europa e no mundo através da internacionalização dos seus produtos e das suas actividades;
6. Captar novas procuras na região através da extensão e alargamento da formação oferecida pela Instituição.

4.4.3. GOVERNO E GESTÃO

O IPT, através da aprovação em Assembleia Estatutária dos novos estatutos, projectou a evolução para um modelo organizacional numa lógica empresarial, que garanta capacidade estratégica e de decisão centralizada, de operacionalização e de responsabilização descentralizadas com autonomia científica e pedagógica. Para tal, pretende adoptar uma estrutura do tipo matricial que cruze produtos e competências, apresentando-se as Escolas como gestoras de produtos, os Departamentos como gestores de carreiras científicas e o Instituto como pólo dinamizador de investigação aplicada e de prestação de serviços, ancorada no desenvolvimento da inovação e do capital humano.

4.4.4 PROCESSO ESTRATÉGICO

Numa altura em que a envolvente social, política e económica nacionais e internacionais do ensino superior sofrem cada vez mais alterações, torna-se fundamental definir uma visão estratégica.

A função estratégica no IPT, está confiada aos órgãos de gestão, sendo o processo de planeamento (plano de actividades) o seu processo mais saliente.

Da leitura do QUAR 3 do IPT e do Plano Estratégico 2007/2013 foram retirados os respectivos objectivos estratégicos:

Objectivo 1: Reformulação do modelo organizacional e reposicionamento da oferta formativa nuclear.

Objectivo 2: Abertura ao exterior, quer no contexto nacional, quer no contexto internacional.

Objectivo 3 Alargamento da área de influência e extensão dos produtos.

Áreas Estratégicas	Eixos Estratégicos
Reformulação do modelo organizacional e reposicionamento da oferta formativa nuclear	1. Reformular o modelo organizacional e de gestão do IPT
	2. Reposicionar a oferta formativa nuclear do IPT
Abertura ao exterior	3. Aprofundar a integração do IPT na envolvente regional de proximidade
	4. Aprofundar a integração do IPT na Rede Nacional de Ensino Superior e no Sistema Nacional de Inovação
Alargamento da área de influência e extensão dos produtos	5. Intensificar a internacionalização
	6. Expandir a oferta formativa, captando novas procuras

Figura 24 - Áreas e Eixos de intervenção - Fonte Plano Estratégico (2007)

³ QUAR Quadro de Avaliação e Responsabilização

QUADRO 3-1:
Objectivos e síntese das linhas de acção (Eixo 1)

	Objectivos	Operacionalização (Linhas de Acção)
Desenvolver uma plataforma de aglutinação estratégica, descentralização operacional e autonomia científica e pedagógica, que articule de forma matricial e efectiva competências e produtos	Evoluir para um modelo de organização que permita maior eficiência na gestão dos recursos, maior eficácia na oferta de produtos e a optimização da gestão financeira do IPT. Implementar um modelo organizacional que, funcionando numa lógica empresarial/holding, garanta capacidade estratégica e de decisão centralizada, responsabilização descentralizada e autonomia científica e pedagógica.	<ol style="list-style-type: none"> 1. Adotar uma estrutura de tipo matricial que cruze produtos e competências. 2. Reflectir na organização a importância da gestão integrada do produto e a operacionalização e responsabilização descentralizadas através da figura do director de produto. 3. Melhorar a capacidade do IPT para definir e comunicar uma estratégia clara e consequente ao nível da investigação e da prestação de serviços. 4. Desenvolver a autonomia científica e pedagógica. 5. Manifestar uma capacidade de liderança na região e integrar nas suas decisões estratégicas contribuições exteriores ao IPT. 6. Optimizar a gestão financeira e diversificar as fontes de financiamento. 7. Implementar métodos de avaliação permanente e processos de acreditação externa. 8. Elaborar uma estratégia de marketing institucional e dos produtos. 9. Estruturar o Gabinete de Estudos e Planeamento .

Figura 25 - Objectivos e síntese das linhas de acção - Fonte Plano Estratégico (2007)

No Eixo 1, linha de acção 7 (Implementar métodos de avaliação permanente e processos de acreditação externa).

Toda a actividade das instituições, e portanto todos os seus produtos, devem continuamente ser objecto de avaliação interna e com frequência externa como caminho indispensável para ver esses mesmo produtos acreditados a nível nacional e/ou internacional. A avaliação e monitorização implicam, necessariamente, a existência de bases de dados actualizadas que permitam gerar a informação necessária à avaliação.

Toda a actividade das instituições, e portanto todos os seus produtos, devem continuamente ser objecto de avaliação interna e com frequência externa como caminho indispensável para ver esses mesmo produtos acreditados a nível nacional e/ou internacional. A avaliação e monitorização implicam, necessariamente, a

existência de bases de dados actualizadas que permitam gerar a informação necessária à avaliação.

Neste sentido torna-se indispensável implementar:

- *A criação de sistemas de informação e análise do feed back dos estudantes da instituição enquanto alunos e depois de terem deixado a instituição pela primeira vez (a manutenção de bases de dados actualizadas sobre os antigos alunos permite não só aferir a empregabilidade associada a cada curso, mas também se revela um instrumento fundamental para a promoção de formação pós-graduada);*
- *A organização sistemática do sistema de avaliação, designadamente externa;*
- *A participação em sistemas de acreditação.*

(Augusto Mateus e Associados, 2007, p. 47)

4.4.5 PRÁTICAS DE AVALIAÇÃO DE QUALIDADE

O Instituto Politécnico de Tomar apenas estabeleceu uma estrutura formal para gestão dos processos internos de avaliação e garantia da qualidade em Setembro de 2007. Esta estrutura, o GAQ (Gabinete de Avaliação e Qualidade), iniciou as suas actividades apenas em Fevereiro de 2008.

4.4.5.1 Procedimentos Internos de Avaliação e Garantia da Qualidade

Cabe o Presidente do IPT decidir sobre a aprovação de novos cursos ou encerramento de cursos em funcionamento, sob proposta do Conselho Científico da Escola a que estão afectos.

Os Conselhos Científicos são responsáveis pela revisão e adequação dos planos de estudo e dos conteúdos programáticos das unidades curriculares.

As Escolas do IPT possuem regulamentos académicos próprios, relativos à inscrição, frequência e avaliação das unidades curriculares. Estes regulamentos estão disponíveis publicamente (nomeadamente nas páginas web do Instituto e Escolas) e são específicos para os cursos de 1º ciclo (Licenciatura), 2º ciclo (Mestrado) e Cursos Tecnológicos (CET). Cada responsável por uma unidade curricular deve, antes do início do seu funcionamento, entregar no respectivo secretariado académico, um programa curricular, indicando os conteúdos programáticos, objectivos, bibliografia recomendada e método de avaliação utilizado. Esta informação é igualmente disponibilizada através das páginas web das Escolas.

O papel dos estudantes e dos parceiros institucionais na avaliação e garantia de qualidade da Instituição decorre da participação nos órgãos de gestão em que se encontram representados. Informalmente, os estudantes são ainda ouvidos em muitos aspectos do funcionamento dos cursos, ao nível dos departamentos, como na elaboração de horários e de calendários de avaliação.

4.5 AVALIAÇÃO EUA

A EUA resultou da fusão entre a Associação das Universidades Europeias (CRE) e a Confederação das Conferências dos Reitores da U.E., ocorrida em 31 de Março de 2001, e representa, simultaneamente, as universidades europeias e os reitores, o que lhe confere um importante papel no seio da comunidade europeia de Ensino Superior.

A missão da EUA é promover o desenvolvimento de um sistema europeu de Ensino Superior e de Investigação coerente, através de um apoio activo e de orientação dos seus membros, enquanto instituições autónomas empenhadas no desenvolvimento da qualidade do ensino ministrado, da aprendizagem e investigação e das contribuições para a sociedade.

A EUA tem como objectivo a criação de uma área de Ensino Superior e Investigação com uma identidade única, baseada nos valores comuns da autonomia das instituições, da educação e da investigação. Visa a compatibilidade das estruturas de Ensino Superior do espaço europeu através de normas reconhecidas e aceites pelos seus membros, de

forma a consolidar o papel do Ensino Superior na Sociedade do Conhecimento, seja em termos de inovação ou disseminação.

4.6 AVALIAÇÃO A3ES

A Agência de Avaliação e Acreditação do Ensino Superior (A3ES), é uma fundação de direito privado, dotada de personalidade jurídica e reconhecida como de utilidade pública, tendo sido criada através do Decreto-Lei n.º 369/2007, de 5 de Novembro tendo em vista a promoção e a garantia da qualidade do ensino superior.

A A3ES tem como missão contribuir para a melhoria da qualidade do ensino superior em Portugal, através da avaliação e acreditação das instituições de ensino superior e dos seus ciclos de estudos.

O objectivo primordial desta Agência é o de proporcionar a melhoria da qualidade do desempenho das instituições de ensino superior, dos seus ciclos de estudos e garantir o cumprimento dos requisitos básicos do seu reconhecimento oficial.

Estes objectivos são prosseguidos, através da avaliação e da acreditação de instituições, dos ciclos de estudos de ensino superior, e por meio destas, da promoção de uma cultura institucional interna de garantia da qualidade.

São objectivos da A3ES:

- Desenvolver a avaliação da qualidade de desempenho das instituições de ensino superior e dos seus ciclos de estudos;
- Concretizar os critérios de avaliação, de modo a obter a tradução dos seus resultados em apreciações qualitativas, bem como definir as consequências da avaliação efectuada para o funcionamento das instituições e dos seus ciclos de estudos;
- Promover a acreditação de ciclos de estudos e instituições, tendo em vista a garantia de cumprimento dos requisitos legais do seu reconhecimento;

- Promover a divulgação fundamentada à sociedade sobre a qualidade do desempenho das instituições de ensino superior;
- Promover a internacionalização do processo de avaliação.

O regime de avaliação e acreditação a desenvolver pela Agência é o constante da Lei n.º 38/2007, de 16 de Agosto.⁴

A entrada no sistema de um novo actor – Agência de Avaliação e Acreditação do Ensino Superior (A3es) - alterou o quadro de competências em matéria de acreditação de ciclos de estudo (Decreto-Lei 369/2007). Com efeito, em finais de 2009, a A3es lançou o primeiro concurso para acreditação de novos ciclos de estudo, cursos para os quais o funcionamento deveria ter início no corrente ano lectivo (2010/2011).

⁴ (Retirado do Site <http://www.a3es.pt/pt/> em Dezembro de 2010)

5 DESENHO METODOLÓGICO

O objectivo deste capítulo, consiste em apresentar um desenho metodológico para o desenvolvimento de um DW numa IES, com um caso de estudo onde se aplica a metodologia proposta.

A abordagem *Top-Down*, preconizada por *Inmon* é centrada nos dados, tem um custo elevado e um tempo de implementação muito alto, mas em compensação fornece uma base de dados integrada e corporativa. Nesta abordagem é definido em primeiro lugar, o modelo de dados corporativo do sistema de *data warehousing*, depois as ferramentas *ETL* e só depois as aplicações de consulta e análise.

Para este autor o *DW* é uma parte do sistema global de *Business Intelligence*. Uma empresa tem um *DW*, e os *DM* obtêm a sua informação do *DW*.

A abordagem *Botton-Up*, seguida por Ralph Kimball, tem uma implementação e retorno rápido, baseando-se na construção de *data marts* independentes, sem a necessidade da definição de uma infra-estrutura corporativa para o *DW* da organização. Cada *DM* tem sua própria área de estágio, assim como processos de extração, transformação e limpeza independentes. A sua característica é a construção de um *DW* incremental a partir do desenvolvimento dos *data marts* independentes. Este autor, define basicamente um projecto de *data warehousing*, como um plano para a migração de dados entre os sistemas operacionais, e um repositório especializado para suporte à decisão, designado por *data warehouse*. Ralph Kimball considera o *DW* como sendo o conglomerado de todos os *DMs* da organização.

A implementação *BUS* é a junção das abordagens *Top-Down* e *Botton-Up*, e é onde se enquadra a abordagem *Microsoft*. É uma implementação que utiliza dimensões e factos padronizados, onde o *DW* lógico é constituído por *data marts* integrados, conseguindo-se assim um *DW* integrado e incremental.

Uma das vantagens desta abordagem é que os primeiros resultados são apresentados rapidamente ao utilizador final, sem perder a visão integrada da organização, e dada a sua flexibilidade, permitir uma rápida absorção de novos requisitos.

A escolha da abordagem de implementação está relacionada com vários factores, como por exemplo a infra-estrutura tecnológica, o âmbito da implementação, o custo associado, o tempo de implementação e a necessidade ou não de um acesso corporativo.

A abordagem escolhida para a implementação do *DW* do IPT foi a em Bus onde se enquadra a abordagem Microsoft, pois além de permitir actuar individualmente sobre cada área de negócio, apresenta resultados rápidos, sem que toda a instituição esteja coberta pelo projecto, e a salvo da introdução de novos sistemas de informação ou, a alterações na orgânica interna dos SI existentes. Esta abordagem vai também permitir que o trabalho já desenvolvido, tenha a intervenção directa de cada um dos colaboradores da área em análise, facilitando a questão da resistência à mudança e consequentemente facilitando a sua aceitação. Em termos tecnológicos a instituição tem também adoptado a utilização de sistemas *Microsoft*, o que facilita a escolha desta metodologia. Finalmente, possibilita também a integração parcial de diferentes fontes de dados, ou seja, os dados de cada base de dados de origem, podem ser adicionados em momentos diferentes.

O método proposto pela *Microsoft* para sistemas de *Data Warehouse*, passa por identificar as componentes do projecto e a execução de um conjunto de actividades.

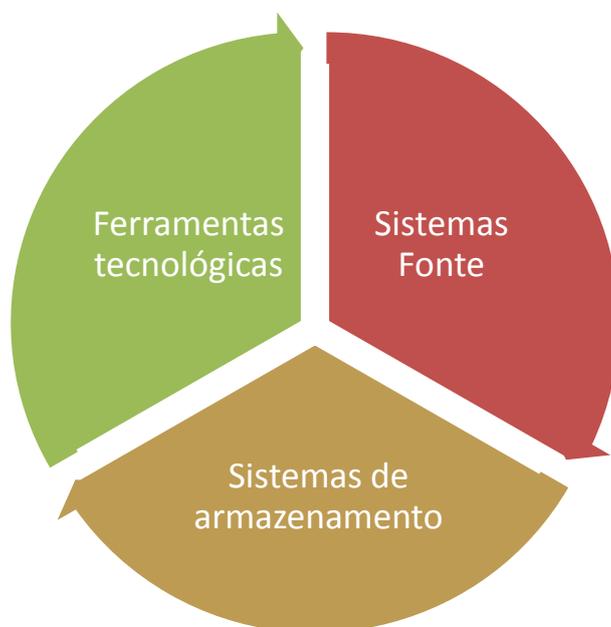


Figura 26 - Componentes do DW - Fonte: Elaboração própria

- Sistemas fonte - levantamento dos sistemas que detêm a informação;

- Sistemas de armazenamento - sistema onde fica alojado o *Data Warehouse* físico;
- Ferramentas aplicacionais de exploração, modelação, desenvolvimento e manutenção.

O método inclui a execução das actividades seguintes:

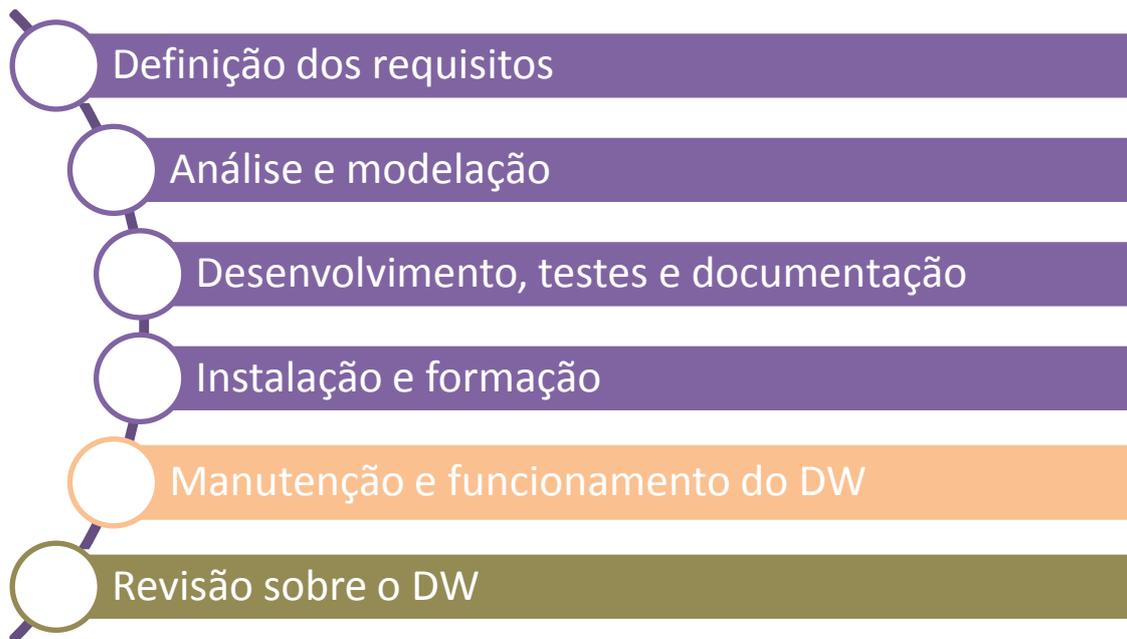


Figura 27 - Actividades do DW - Fonte: Elaboração própria

5.1 SISTEMAS FONTE EXISTENTES

Nesta componente, efectuou-se o levantamento dos sistemas de informação e das aplicações informáticas, existentes na organização. Assim foi criada uma ficha tipo de levantamento de sistemas com a informação seguinte:

- Objectivo – qual o objectivo do sistema e quais os principais problemas que a aplicação pretende resolver;
- Tipo de sistema – classificação do tipo de sistema de acordo com o nível da organização em que é utilizado;
- Funcionalidades – descrição das principais funcionalidades da aplicação agrupadas em módulos quando necessário;

- Dados de exportação para o *Data Warehouse* – Identificação dos dados que vão ser exportados para o *Data Warehouse*.

Exemplo da ficha por Aplicação tipo:

AP1 – Sistema de gestão de alunos

OBJECTIVO

A área Académica do SiGES é composta por aplicações que oferecem uma solução para a gestão das tarefas académicas associadas às comunidades discente e docente das Instituições de Ensino Superior. Gerir as inscrições dos alunos no curso e nas disciplinas, incluindo os planos de estudos. Registrar as candidaturas, que são vistas como pré-inscrições nos cursos.

TIPO DE SISTEMA

TPS - Transaction Processing Systems (Sistemas de Processamento de Transacções).

FUNCIONALIDADES

Módulo CSE – Controlo de Sistema de Ensino

A aplicação permite o controlo de várias Escolas ou Pólos, sendo possível gerir planos inter-curriculares, troncos ou áreas comuns, sistemas de créditos, sistemas mistos e uma total abrangência de regras de transição de ano e de cálculos de fim de curso.

Módulo CXA – Controlo e Gestão de Tesouraria

Este módulo permite gerir a conta corrente de alunos, candidatos, funcionários e de outras entidades, emitindo facturas, recibos com ou sem parcelamento, créditos e débitos. Permite o cálculo de propinas e planos de multas.

Módulo SIA Sistema de Inscrições Automáticas

Este módulo permite gerir todas as inscrições dos alunos e inclui o SIA Óptico que permite a utilização de códigos de barras nas inscrições.

DADOS PARA INTEGRAÇÃO COM O *DATA WAREHOUSE*

Campos necessários
Identificação dos Campos que vão para o Data Warehouse
Nome do aluno, Numero do Aluno, Morada do aluno, código postal do aluno, concelho da morada, freguesia da morada, concelho de nascimento, freguesia de nascimento, número fiscal, numero de conta caixa.
Denominação do curso, grau académico do curso, ano de criação, código do curso, curso origem, curso destino, decreto lei do curso, data revisão do curso, ciclo de estudos, ramos, perfis.
Código da unidade curricular, nome da unidade curricular, curso, horas de trabalho.
Nº de conta, nº de aluno, pagamento, data pagamento, tipo de pagamento, data do documento.

AP2 – Sistema de gestão Financeira

OBJECTIVO

O SIAG (Sistema Integrado de Apoio à Gestão) é um sistema que cobre áreas que vão desde a Gestão Estratégica à Gestão Operacional numa organização.

TIPO DE SISTEMA

TPS - Transaction Processing Systems (Sistemas de Processamento de Transacções).

FUNCIONALIDADES

Módulo Gestor – Controlo de Sistema de Financeiro\

A aplicação permite a gestão dos recursos financeiros e a correspondente execução orçamental.

Módulo Fact – Facturação de Bens e Serviços

Visa proporcionar a facturação de bens e serviços.

Módulo RH+ - Gestão de Recursos Humanos

Este módulo visa proporcionar a gestão e administração dos respectivos Recursos Humanos.

Sub-Módulo RH+ - Gestão de Pessoal

Permite efectuar a gestão do quadro de pessoal de cada serviço ou Organismo bem como registar e manter actualizado o cadastro de cada funcionário ou agente. Contém Quadro de Pessoal, Cadastro, Assiduidade, Lista de Antiguidade.

Sub-Módulo RH+ - Vencimentos

Permite o cálculo das remunerações dos funcionários e agentes. Para além do tratamento das carreiras e categorias, de acordo com o estatuto remuneratório dos funcionários e agentes da Administração Pública, são observadas com rigor todas as regras estabelecidas para o cálculo dos abonos e dos descontos. Contém Tabelas Salariais, Cálculo das Remunerações, Centros de Controlo e Apuramento, Mapas e Apuramentos, BALANÇO SOCIAL.

Módulo Spid - Sistema de Processamento de Deslocações

Este módulo visa proporcionar a contabilização e processamento de Deslocações em serviço oficial, designadamente Ajudas de Custo e Transportes em Território Nacional e Estrangeiro.

Módulo SIME- Sistema de Inventário e Património

Este módulo visa proporcionar o registo, inventariar a gestão e cadastro de todos os bens da instituição.

DADOS PARA INTEGRAÇÃO COM O *DATA WAREHOUSE*

Campos necessários
Identificação dos Campos que vão para o <i>Data Warehouse</i>
Nome do docente, Numero do docente, Morada do docente, código postal, concelho da morada, freguesia da morada, concelho de nascimento, freguesia de nascimento, número fiscal, unidades curriculares que lecciona, anos lectivos, cursos, Grau, área científica, regime de tempo,

AP3 – Sistema de Bolsas Serviços de Acção Social

OBJECTIVO

O INFOS é um sistema que cobre a área da gestão de bolsas de estudo.

TIPO DE SISTEMA

TPS - Transaction Processing Systems (Sistemas de Processamento de Transacções).

FUNCIONALIDADES

Módulo Bolsas de Estudo

A aplicação permite o registo e armazenamento dos dados dos alunos indispensáveis à avaliação da sua candidatura a bolsa de estudo e à prestação das informações complementares.

Determinar com base nos dados recolhidos quais os alunos que cumprem os requisitos para a atribuição de bolsa e o respectivo valor.

DADOS PARA INTEGRAÇÃO COM O *DATA WAREHOUSE*

Campos necessários
Identificação dos Campos que vão para o <i>Data Warehouse</i>
Nome do aluno, Numero do Aluno, número fiscal, data de pedido de bolsa, tipo de bolsa concedida, ano lectivo, valor da bolsa.

AP4 – Sistema Integrado de gestão de bibliotecas

OBJECTIVO

Sistema de gestão de monografias, catalogação de periódicos, gestão de kardex aquisições de periódicos e não periódicos, fornecedores, centros de custo e orçamentos, base de dados de leitores e respectivo perfil, empréstimos, filas de espera, etc..

TIPO DE SISTEMA

TPS - Transaction Processing Systems (Sistemas de Processamento de Transacções).

MACRO FUNCIONALIDADES

Módulo Gestão de Empréstimos

A aplicação permite a gestão dos empréstimos de todos os livros.

Módulo Gestão de Arquivos

Visa proporcionar toda a gestão de arquivos/documentação produzida na instituição.

DADOS PARA INTEGRAÇÃO COM O *DATA WAREHOUSE*

Campos necessários
Identificação dos Campos que vão para o <i>Data Warehouse</i>
Nome do utilizador, Numero do utilizador, tipo de utilizador (Aluno, Docente, não docente), Área temática do livro, Nome do livro, nº de pedidos.

5.2 SISTEMA DE ARMAZENAMENTO

Corresponde ao repositório físico final de dados e onde, com a finalização da modelação dimensional, deve ser criada a base de dados do *DW*.

Existe no mercado um conjunto de sistemas que podem ser candidatos a receber o *DW*, contudo devem prever o armazenamento dos registos em ambiente multidimensional, esquemas em estrela, com informação agregada em estruturas do *OLAP*.

As ferramentas *OLAP* são classificadas de acordo com a estratégia de armazenamento em:

- *Relational OLAP (ROLAP)* - Usam um sistema *DBMS* relacional para guardar e gerir os dados do *DW*, e *middleware OLAP* para suportar as *queries*. Exemplo: *DSS MicroStrategy*;
- *Multidimensional OLAP (MOLAP)* - Suportam visões multi-dimensionais dos dados, através de motores multidimensionais, suportados baseados em *arrays*. Providencia uma rápida indexação dos dados pré sumarizados. Exemplo: *Essbase da Arbor*;
- *Hybrid OLAP (HOLAP)* - Combinam as tecnologias *ROLAP* e *MOLAP* e permitem, por exemplo, que grandes volumes de dados sejam guardados em estruturas relacionais, enquanto que algumas agregações são usadas em cubos, por forma a aumentar a rapidez da análise dos seus dados;
- Servidores *SQL* especializados - Algumas empresas implementam servidores *SQL* especializados que providenciam uma linguagem de interrogação das *BDs* que suporta *SQL* sobre modelos dimensionais em modo *read-only*.

A infra-estrutura tecnológica sobre a qual foi implementado o *DW*, assenta na tecnologia *Microsoft* tendo por sistema operativo o *Windows Server 2008*, sistema operativo para servidores de aplicações da *Microsoft* e como Sistema de Gestão de Base de dados o *Microsoft SQL Server 2008*, que permitiram a criação, implementação, análise e construção das tabelas da base de dados.

A escolha da plataforma *Microsoft* deve-se à disponibilidade de obtenção da mesma e por já existir a funcionar no Centro de Informática do IPT.

5.3 FERRAMENTAS DE EXPLORAÇÃO, MODELAÇÃO E EXPLORAÇÃO

Para a construção deste projecto, são propostas algumas ferramentas, a sua selecção não teve em linha de conta quais as melhores ferramentas ou as que possuíam melhor desempenho, mas sim ao conhecimento das pessoas envolvidas neste projecto. Para o seu desenvolvimento propõe-se a utilização do seguinte:

- a) Ferramentas de Exploração - Estas ferramentas devem, também, ser identificadas por tipo de utilizadores como, por exemplo: a1) os que só lêem relatórios; a2) os que produzem relatórios; a3) analistas de negócio ou super-utilizadores; a4) gestores (executivos);
 - a. *Oracle Discover*: é uma poderosa ferramenta de extracção de dados voltada para *Data Warehouse* e data-mart (ou *OLTP*) sem programação que permite a publicação de dados directamente na *WEB*, ou exportação para Excel;
 - b. *Oracle User Edition* que é uma ferramenta de extracção de dados que acaba com a necessidade de usar complexos comandos *SQL*, também possui a capacidade de executar *drillanywhere* e de criar *reports*;
 - c. Microsoft Office Excel, pode ser utilizado para a fase de limpeza dos dados no processo de *ETL* do *DW* e para consulta de dados e relatórios.

- b) Ferramentas de Modelação e desenvolvimento – consistem nas aplicações informáticas utilizadas para construir os ambientes de exploração;
 - a. Linguagem de Programação *SQL*: Ferramenta utilizada para a criação das tabelas em formato texto. Estas tabelas serão utilizadas posteriormente no carregamento do *Data Mart*;

- b. *Procedures Oracle*: Ferramenta utilizada para carregar as tabelas de dimensão e de factos do modelo dimensional, a partir das tabelas no formato texto;
- c. *Microsoft Visio*, utilizado inicialmente para criar o modelo dimensional do DW.

c) Ferramentas de manutenção: são aplicações informáticas utilizadas para gerir o desempenho, segurança e integridade do Sistema de *Data Warehouse*.

5.4 ACTIVIDADES PARA A IMPLEMENTAÇÃO DO SISTEMA DE *DATA WAREHOUSE*

Conforme descrito anteriormente, o método proposto inclui a execução de oito actividades, onde quatro fazem parte do grupo da implementação do Sistema de *Data Warehouse*, duas referem-se ao grupo das tarefas de revisão e outras duas referem-se ao grupo de manutenção. Reforça-se que este processo é iterativo, isto é, o resultado das actividades de revisão e de manutenção fornecerão indicações ou recomendações para as próximas actividades de construção.

5.4.1 DEFINIÇÃO DE REQUISITOS

Uma definição bastante simples para requisitos é dada por (Macaulay, 1996), segundo o qual, requisito é “*simplesmente algo de que o cliente necessita*”.

Um requisito é uma condição que tem que se verificar para atingir os objectivos do sistema, o termo requisito é definido como:

1. Uma condição ou capacidade que um utilizador necessita para resolver um problema ou atingir um objectivo;
2. Uma condição ou capacidade que deve ser satisfeita ou possuída por um sistema ou uma componente de um sistema, para satisfazer um contrato, um *standart*, uma especificação ou qualquer outro documento, formalmente, imposto;

3. Uma representação documentada de uma condição ou capacidade, tal com em 1 ou 2 ((Filomena Castro Lopes, 2009).

Tanto Kimball como Inmon, recomendam entrevistar uma comunidade de utilizadores bem definidos, basicamente aqueles que têm maior poder nas tomadas de decisões ou com amplo conhecimento sobre o negócio ou sobre o *Data Mart* em estudo.

Segundo Kimball, desenvolver um DW é uma questão de unir as necessidades dos seus utilizadores, com a realidade dos dados disponíveis. Este autor aponta também um conjunto de três pontos fundamentais no projecto da estrutura de um DW, que constituem definições a serem feitas, as quais correspondem às etapas do projecto (Kimball R. , 1996).

1. Escolha de um processo do negócio para modelar - Um processo do negócio é uma operação importante da sua organização suportada por algum tipo de sistema fonte (transaccional) de onde é possível recolher os dados para o futuro *Data Warehouse*.
2. Escolha do grão do processo do negócio - O grão é o nível fundamental atômico de dados que representará esse processo na tabela de factos. A granularidade é a medida do detalhe dos dados. Os dados transaccionais têm o nível de granularidade mais baixo, porque têm a informação mais detalhada possível. Este nível tem também a designação de atómico.
3. Escolha das dimensões que serão aplicadas a cada registo da tabela de factos - Para cada dimensão escolhida, temos de descrever todos os diferentes atributos de dimensão (campos) que preenchem cada tabela dimensional.
4. Escolha dos factos a medir que irão povoar cada registo da tabela de factos.

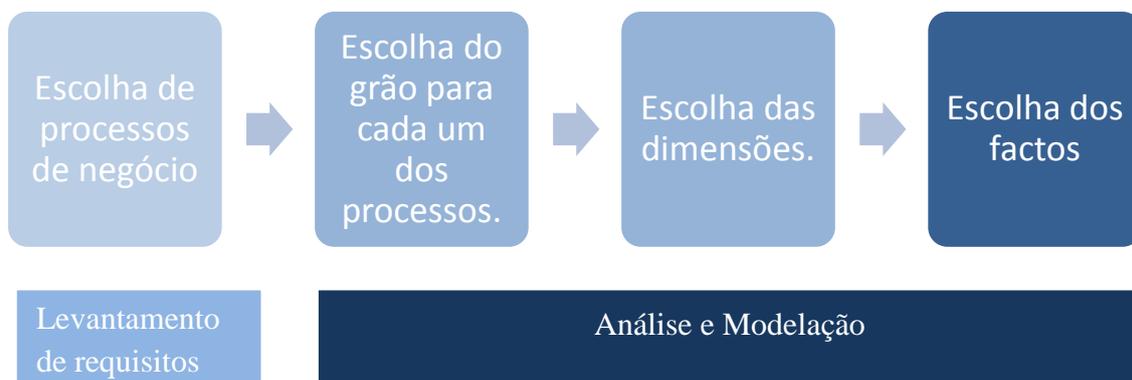


Figura 28 - Esquema para Levantamento requisitos e análise e modelação - Fonte: Elaboração própria

Independentemente das fontes, o objectivo do levantamento de requisitos é interiorizar o conhecimento sobre o sistema para melhor o entender. Existem várias técnicas (análise de documentação, entrevistas, questionários e observação directa são as mais comuns) que podem ser usadas para atingir esse objectivo, permitindo e facilitando o diálogo com os utilizadores e, por conseguinte, possibilitando a escolha da informação necessária para construir a imagem do sistema em estudo.

Para o desenvolvimento desta fase do projecto, foram consultados os guias de avaliação de cursos do EUA e da A3e's de forma a podermos identificar que informação é necessária recolher, foram efectuadas consultas aos utilizadores do Gabinete de Avaliação e Qualidade para, também aqui, aferir as necessidades de informações para o processo de avaliação, juntamente com a recolha dos pedidos avulsos solicitados ao Centro de Informática pelos diversos utilizadores.

De entre estas fontes escolheu-se um conjunto de questões para o desenvolvimento do DW no IPT, sendo condição obrigatória pertencerem ao Guia de Avaliação da A3E's que se encontra no Anexo A deste relatório.

- Caracterização dos estudantes inscritos no ciclo de estudos, incluindo o seu género, idade, região de proveniência e origem socioeconómica (escolaridade e situação profissional dos pais (Questão 5.1.1 do Guião A3e's).
- Procura do ciclo de estudos por parte dos potenciais estudantes nos últimos 3 anos (Questão 5.1.2 do Guião A3e's).
- Comparação do sucesso escolar nas diferentes áreas científicas e respectivas unidades curriculares (Questão 7.1.2 do Guião A3e's).
- Percentagem de diplomados que obtiveram emprego em sectores de actividade relacionados com a área do ciclo de estudos (Questão 7.1.4 do Guião A3e's).
- Número de publicações, do corpo docente do ciclo de estudos em revistas internacionais com revisão por pares, nos últimos 3 anos e na área do ciclo de estudos (Questão 7.2.2 do Guião A3e's).
- Actividades de desenvolvimento tecnológico e artístico, prestação de serviços à comunidade e formação avançada (Questão 7.3.1 do Guião A3e's).

- Percentagem dos docentes que mantêm a sua ligação ao ciclo de estudos por um período superior a três anos (Questão 4.1.12 do Guião A3e's).

5.4.2 ANÁLISE E MODELAÇÃO

Foram, assim, identificados na actividade anterior um conjunto de processos que passam a ser descritos e modelados conforme as fichas seguintes.

Nestas fichas propõe-se que os processos sejam codificados, descritos, identificada a granularidade, a tabela de factos, as dimensões, o exemplo da informação que irá ser disponibilizada ao utilizador, e a respectiva estrela do processo.

Estudantes em números

Processo NAC Numero de inscrições\matriculas por aluno

Descrição do Processo: Este processo de negócio visa saber o número de alunos que ingressaram nos cursos das escolas do IPT por tipo de ingresso (Contingente Geral,) e ano lectivo.

Granularidade: A granularidade usada na modelação deste processo de negócio é anual/sempre que hajam matriculas/inscrições.

Tabela de Factos: A tabela de factos receberá uma linha, para inscrição/matricula do aluno num curso. Os factos de interesse para este modelo são o número de alunos inscritos/matriculados por curso, ano lectivo, tipo de ingresso, e 1º Ano e 1º Ano primeira vez.

A tabela de factos Matriculas Alunos contem informação sobre o número de inscrições do aluno num curso numa data.

Dimensões: As dimensões que farão parte são: Alunos, Cursos e Data.

Admissões devolve factos que podem ser dissecados por Escola, Curso, Tipo de Candidatura, Matricula cujo intervalo temporal serão os anos lectivos. Cada processo normalmente dá origem a uma tabela de factos.

Ano Lectivo	Matriculas					
	1º Ano	1º Ano, 1.º vez	Contingente Geral	M23	Outro tipo	Total
IPT Total	1376	1022	570	192	279	1041
Curso A	74	46	34	4	7	45
Programa CSE	a)	b)	Código de ingresso			

Tabela 9- Tabela matriculas por Ciclo/Ano Lectivo

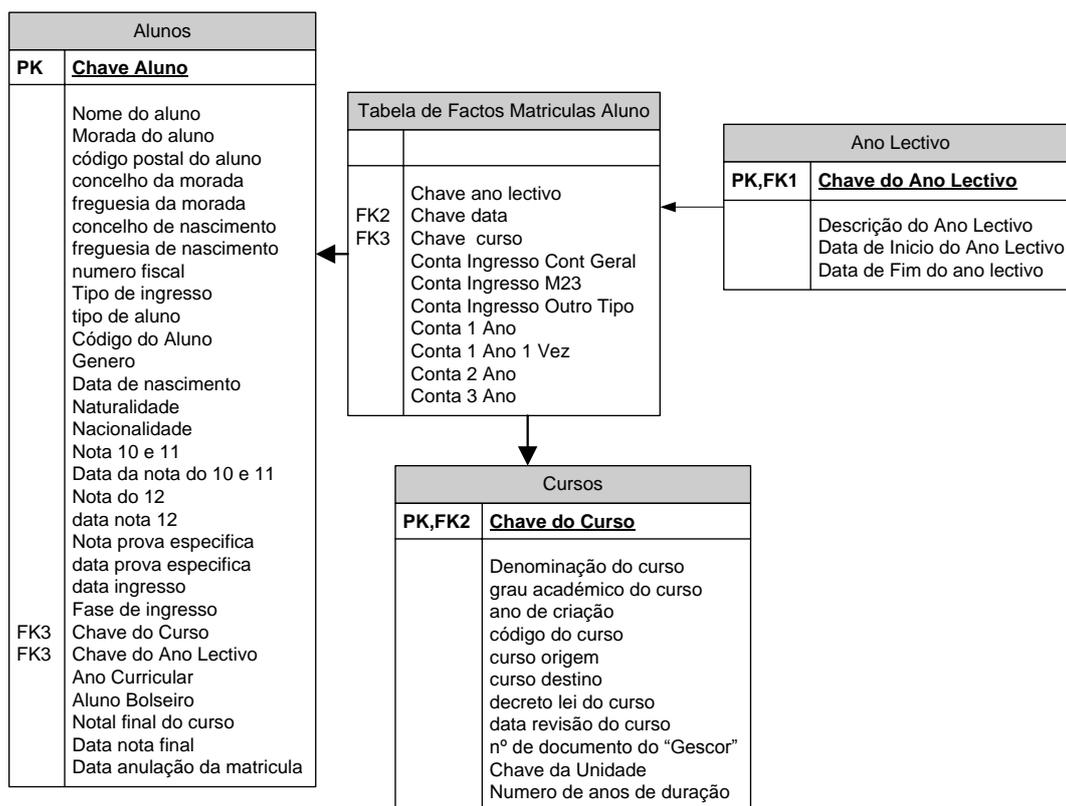


Figura 29 - Esquema Estrela Matriculas alunos - Fonte: Elaboração própria

Processo NAF (Numero de alunos por Fase) De referir que apenas faz sentido para os cursos que conferem grau de licenciatura.

Descrição do Processo: Este processo de negócio visa saber o número de alunos que ingressaram nos cursos das escolas do IPT por Fase (Fase 1, 2 e 3) e ano lectivo.

Granularidade: A granularidade usada na modelação deste processo de negócio é anual/sempre que se realizem matriculas por cada uma das Fases 1, 2 e 3.

Tabela de Factos: A tabela de factos receberá uma linha e soma por curso os alunos matriculados por fase.

A tabela de factos ContaFases contém informação sobre o número total de matrículas por fase, por curso numa data/ano lectivo.

Dimensões: As dimensões que farão parte são: Cursos e Data.

Cursos	1.ª fase							2.ª fase					3.ª fase				
	Candidaturas	Matriculas	Vagas	Candidaturas	Colocados	Nota do 1º Colocado	Nota do último colocado	Vagas	Candidaturas	Colocados	Nota do 1º Colocado	Nota do último colocado	Vagas	Candidaturas	Colocados	Nota do 1º Colocado	Nota do último colocado
IPT Total	2825	865	750	1413	405	xxxx	xxxx	474	1058	291	xxx	xxx					
Exemplo Curso 1	83	37	45	62	26	170,90	119,10	21	19	11	143,10	118,00					
Exemplo Curso 2	148	39	30	116	30	159,80	122,90	8	30	10	150,80	111,30					
Exemplo Curso X	xxx	xx	xx	xx	xx	xxx	xxx	xx	xx	xx	xx	xxx					

Tabela 10 - Candidaturas, Matrículas, Vagas e Colocações.

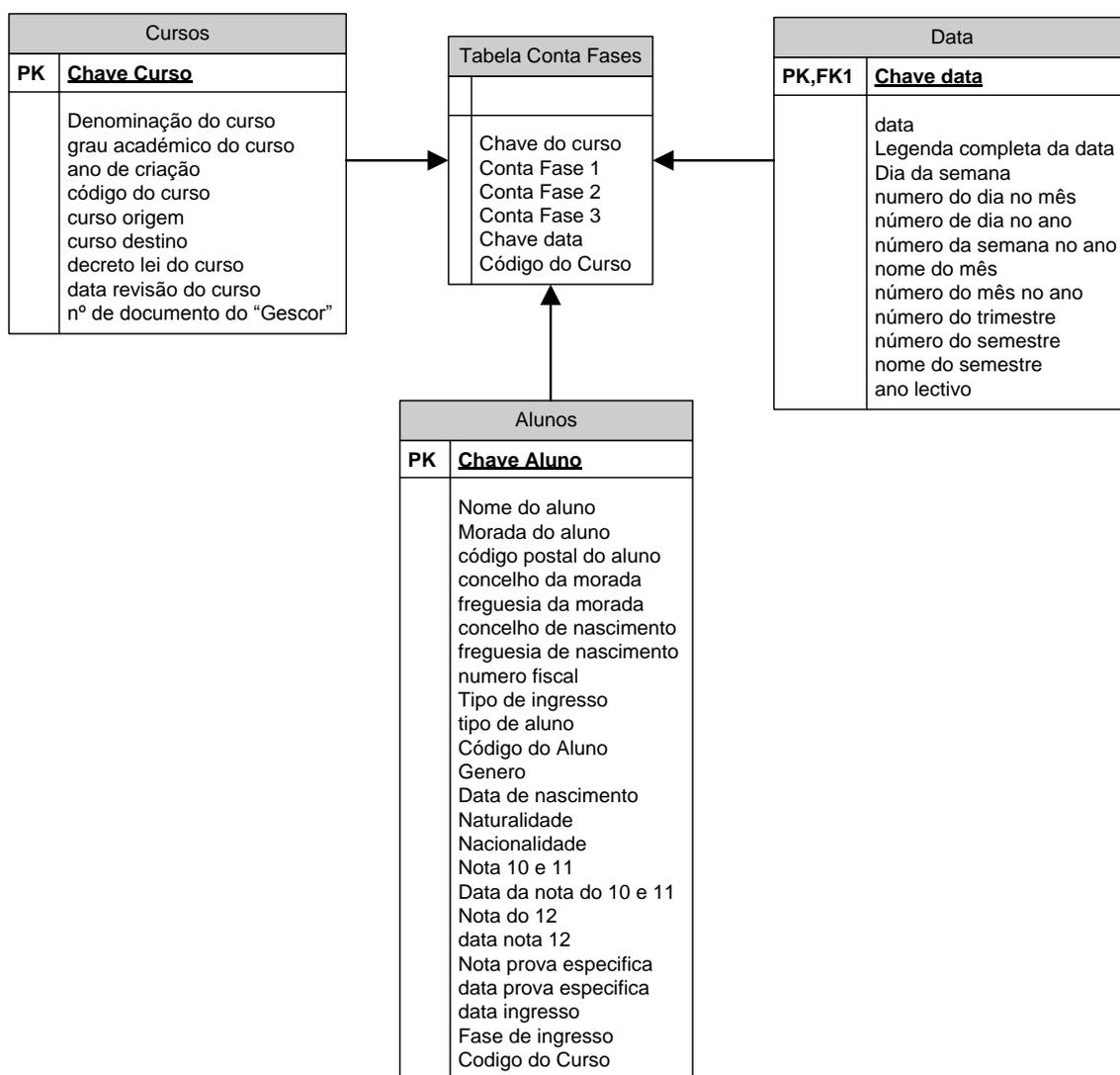


Figura 30 - Esquema Estrela Fases - Fonte: Elaboração própria

Processo NCC (Numero de Candidatos por Curso)

Descrição do Processo: Este processo de negócio visa saber o número de candidatos que pretendem ingressar por ano lectivo nos cursos das escolas do IP.

Granularidade: A granularidade usada na modelação deste processo de negócio é sempre que se realizem candidaturas para os cursos IPT.

Tabela de Factos: A tabela de factos receberá uma linha e soma por curso os candidatos por ano lectivo curso.

A tabela de factos Conta Candidatos contem informação sobre o número total de candidatos por curso/ano lectivo.

Dimensões: As dimensões que farão parte são: Cursos e Ano Lectivo.

	2005/2006	2006/2007	2007/2008	2008/2009
ESTT	1099	1187	1422	1649
ESGT	751	673	1191	1050

Tabela 11 - Evolução do número de candidaturas

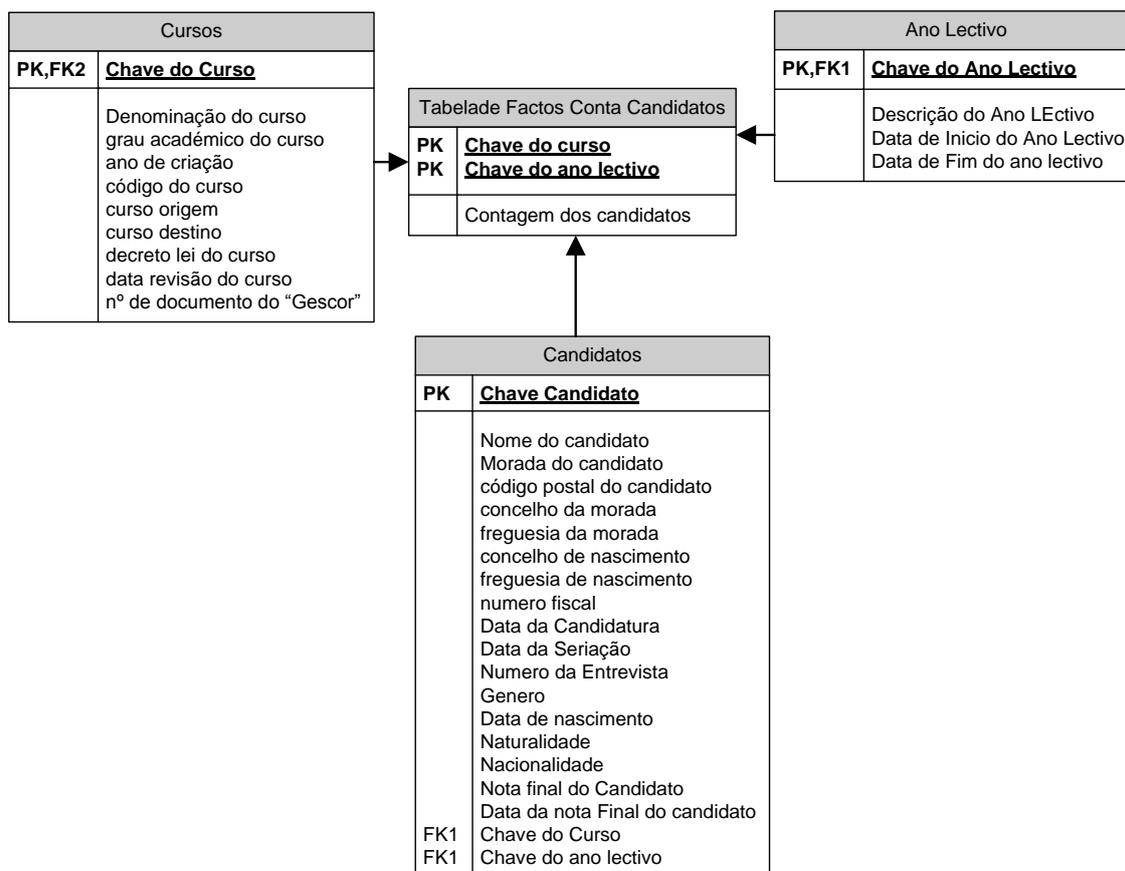


Figura 31 - Esquema Estrela Candidaturas - Fonte: Elaboração própria

Características, Distribuição por Género

Processo NCC (Numero de Alunos por Género)

Descrição do Processo: Este processo de negócio visa saber o número de estudantes distribuídos por Género/Curso.

Granularidade: A granularidade usada na modelação deste processo de negócio é sempre que se realizem Matriculas em cursos.

Tabela de Factos: A tabela de factos receberá uma linha e soma o total de alunos distribuídos por Género, Curso e Ano Lectivo.

A tabela de factos Conta Géneros contém informação sobre o número total de alunos por distribuídos por Género, Curso e Ano Lectivo.

Dimensões: As dimensões que farão parte são: Cursos e Ano Lectivo.

Nota: Correspondência no A3e's: 5.1.1.1.

	Masculino	% Masculino	Feminino	% Feminino	Total
ESTT	910	62%	565	38%	1475
Curso A	XX	XX%	XX	XX%	XY
Curso B	XX	XX%	XX	XX%	XY
ESGT	440	40%	663	60%	1103
ESTA	293	64%	168	36%	461
IPT Total	1643	54%	1396	46%	3039

Tabela 12 - Número de Estudantes do 1º Ciclo Género

	Ano Lectivo N					Ano Lectivo N+1				
	M	% M	F	% F	Total	M	% M	F	% F	Total
ESTT	975	55%	812	45%	1787	966	55%	779	45%	1745
ESGT	352	35%	648	65%	1000	371	36%	647	64%	1018
ESTA	228	55%	188	45%	416	300	61%	189	39%	489

Tabela 13 - Evolução do Número de Estudantes

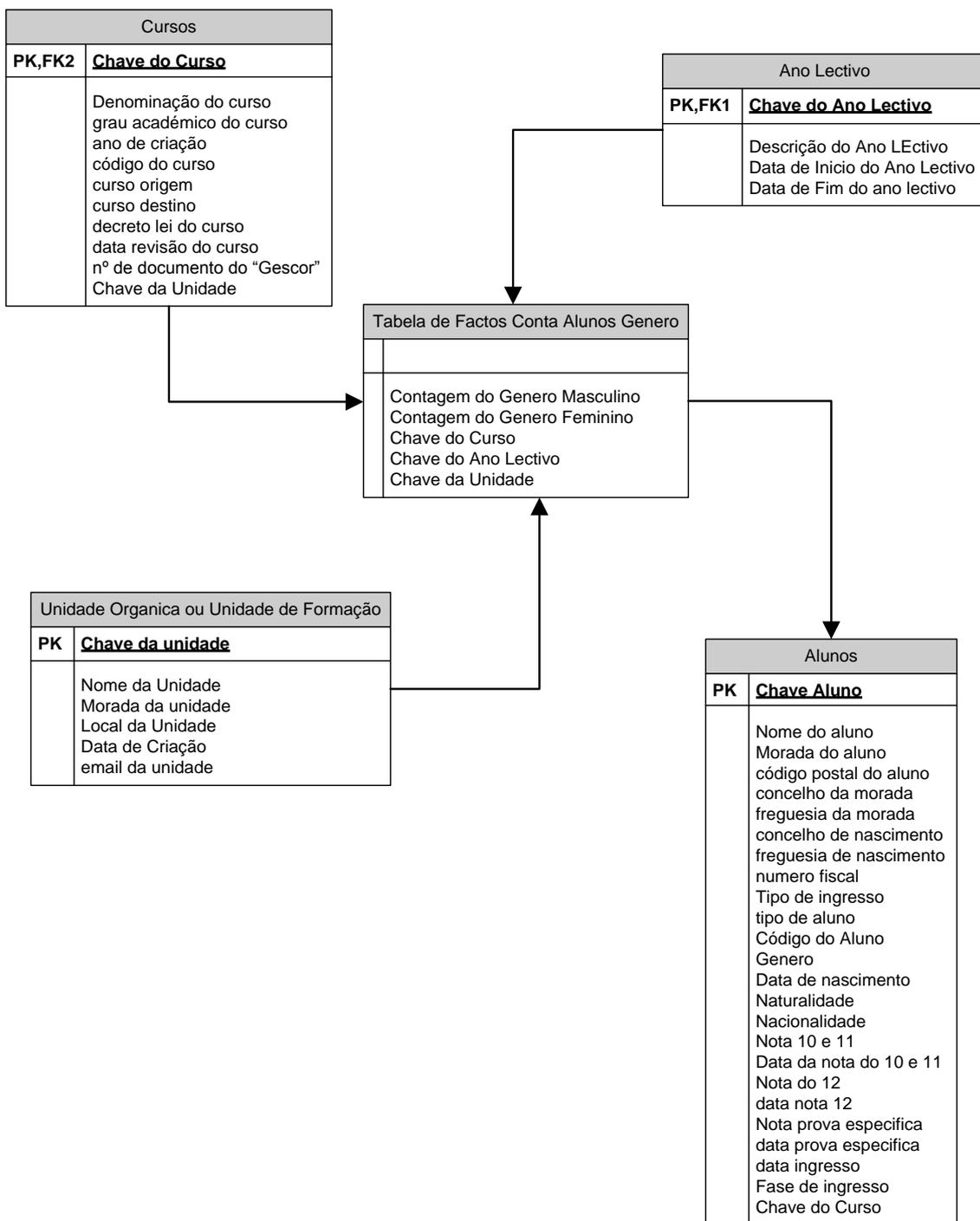


Figura 32 - Esquema Estrela Conta alunos - Fonte: Elaboração própria

Estudantes em números - Faixa etária

Processo NFE (Numero de Alunos por Faixa Etária)

Descrição do Processo: Este processo de negócio visa saber o número de estudantes distribuídos por Faixa Etária/Curso/Escola.

Granularidade: A granularidade usada na modelação deste processo de negócio é sempre que se realizem Matrículas em cursos.

Tabela de Factos: A tabela de factos receberá uma linha e soma o total de alunos distribuídos por idades < que 20 anos, entre 20 e 23, entre 24 e 27, 28 ou mais.

A tabela de factos recebe as idades dos alunos, e contém informação sobre o número total de alunos distribuídos por idades, Curso e Ano Lectivo.

Dimensões: As dimensões que farão parte são: Cursos e Ano Lectivo.

A caracterização dos alunos por idade tem normalmente como referências de idade, < que 20 anos, entre 20 e 23, entre 24 e 27, 28 ou mais para a A3e's.

Para a EUA tem como referência de idade, < que 20 anos, entre 20 e 22, entre 23 e 30, e mais de 30.

Nota: Correspondência no A3e's: 5.1.1.2.

	<20	% <20	20 a 22	% 20 a 22	23 a 30	% 23 a 30	>30	%>30	Total
ESTT	293	20%	453	31%	431	29%	298	20%	1475
ESGT	157	14%	307	28%	328	30%	311	28%	1103
ESTA	67	15%	138	30%	171	37%	85	18%	461
IPT Total	517	17%	898	30%	930	31%	694	23%	3039
	<20		20 a 23		24 a 27		>28		

Tabela 14 - Numero de Estudantes por Faixa Etária

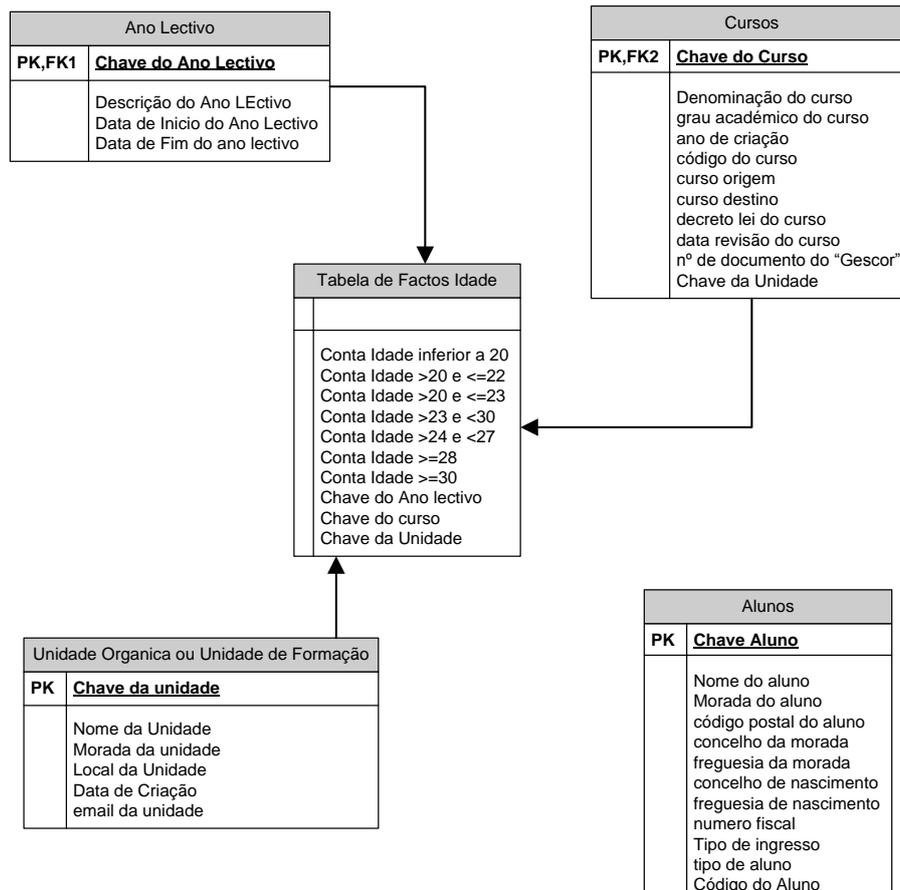


Figura 33 - Esquema Estrela Idades alunos - Fonte: Elaboração própria

Processo NFE (Numero de Alunos por Nacionalidade)

Descrição do Processo: Este processo de negócio visa saber o número de estudantes distribuídos por Nacionalidade/Curso/Escola.

Granularidade: A granularidade usada na modelação deste processo de negócio é sempre que se realizem Matriculas em cursos.

Tabela de Factos: A tabela de factos receberá uma linha e soma o total de alunos distribuídos por Nacionalidade, Curso e Ano Lectivo.

A tabela de factos Conta Nacionalidades contem informação sobre o número total de alunos por distribuídos por Nacionalidade, Curso e Ano Lectivo.

Dimensões: As dimensões que farão parte são: Cursos, Ano Lectivo, Nacionalidades.

Nota: Correspondência no A3e's: 5.1.1.

	Portugal	%	Estrangeiro								Total
			Europa	%	África	%	América	%	Ásia	%	
ESTT	1435	97%	3	0,20%	28	1,90%	9	0,61%	0	0	1475
ESGT	1061	96%	2	0,18%	37	3,35%	3	0,27%	0	0	1103
ESTA	442	96%	0	0,00%	16	3,47%	3	0,65%	0	0	461
IPT Total	2938	97%	5	0,16%	81	2,67%	15	0,49%	0	0	3039

Tabela 15 - Numero de Estudantes por Nacionalidade

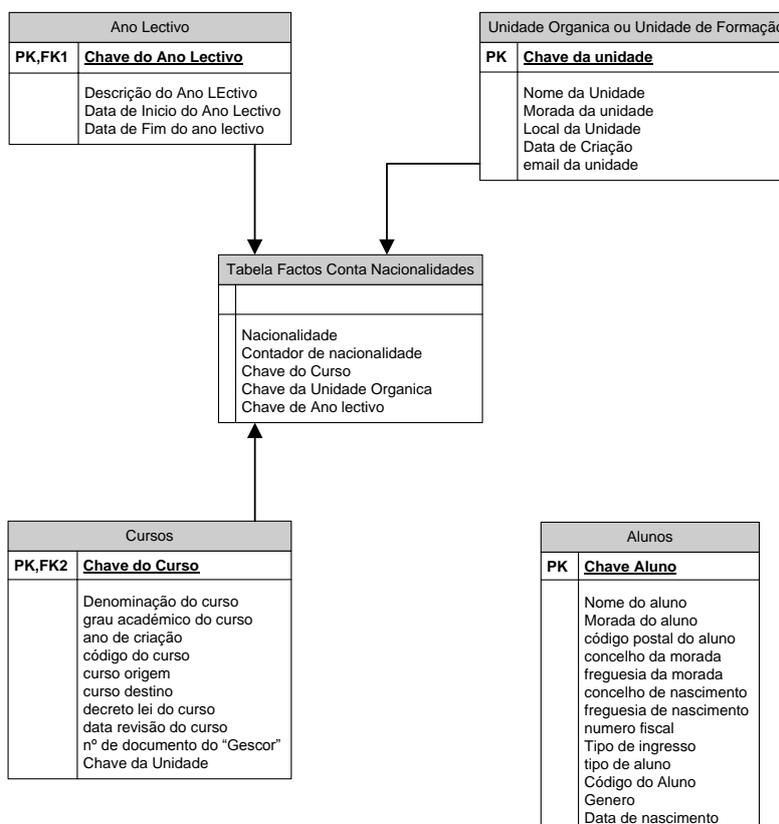


Figura 34 - Esquema Estrela Nacionalidades alunos - Fonte: Elaboração própria

Distribuição por Região de Proveniência Processo NARP (Numero de Alunos por Região de Proveniência)

Descrição do Processo: Este processo de negócio visa saber o número de alunos distribuídos por Região de Proveniência/Curso/Escola.

Granularidade: A granularidade usada na modelação deste processo de negócio é sempre que se realizem Matriculas em cursos.

Tabela de Factos: A tabela de factos receberá uma linha e soma o total de alunos distribuídos por Proveniências Regiões, Curso e Ano Lectivo.

A tabela de factos Conta Proveniências contem informação sobre o número total de alunos por distribuídos por Proveniências, Curso e Ano Lectivo.

Dimensões: As dimensões que farão parte são: Cursos, Ano Lectivo, Proveniências.

Correspondência no A3e's: 5.1.1.3.

	Regiões								Total
	RegA	%	RegB	%	RegC	%	RegD	%	
ESTT									
ESGT									
ESTA									
IPT Total									

Tabela 16 - Numero de Estudantes por Regiões

Distribuição por Ano Curricular

Processo NFE (Numero de Alunos por Ano Curricular)

Descrição do Processo: Este processo de negócio visa saber o número de estudantes distribuídos por Ano Curricular/Curso/Escola.

Granularidade: A granularidade usada na modelação deste processo de negócio é sempre que se realizem Inicio/Fim Ano curricular por cursos/Escolas.

Tabela de Factos: A tabela de factos receberá uma linha e soma o total de alunos distribuídos por Nacionalidade, Curso e Ano Lectivo.

A tabela de factos Conta Alunos Ano Curricular contem informação sobre o número total de alunos por distribuídos por Ano Curricular, Curso e Ano Lectivo.

Dimensões: As dimensões que farão parte são: Cursos, Unidade Orgânica e Ano Lectivo.

	Ano Lectivo 2010/2011						Total
	1.º Ano		2.º Ano		3.º Ano		
	Número	%	Número	%	Número	%	
ESTT	692	47%	398	27%	385	26%	1475
Curso A							
ESGT	464	42%	335	30%	304	28%	1103
Curso B							
ESTA	209	45%	115	25%	137	30%	461
Curso C							
IPT Total	1365	45%	848	28%	826	27%	3039

Tabela 17 - Numero de Estudantes por Escola/Anos Curriculares num Ano Lectivo

Distribuição por tipo de Aluno (Ordinário, TEstudante, ...)

Processo NAT (Numero de Alunos por tipo de aluno)

Descrição do Processo: Este processo de negócio visa saber o número de estudantes por tipo de aluno distribuídos por Ano Curricular/Curso/Escola.

Granularidade: A granularidade usada na modelação deste processo de negócio é Início/Fim Ano curricular por cursos.

Tabela de Factos: A tabela de factos receberá uma linha e soma o total de alunos por tipo distribuídos por Ano Curricular, Curso e Ano Lectivo.

A tabela de factos Conta Alunos Ano Curricular contem informação sobre o número total de alunos por distribuídos por Ano Curricular, Curso e Ano Lectivo.

Dimensões: As dimensões que farão parte são: Cursos, Tipo de alunos, Unidade Orgânica e Ano Lectivo.

	Anos curriculares									Total
	1.º Ano			2.º Ano			3.º Ano			
	Ord	TE	Outro	Ord	TE	Outro	Ord	TE	Outro	
ESTT										1475
Curso A										
ESGT										
Curso B										
ESTA										
Curso C										

Tabela 18 - Estudantes por tipo

Sucesso Escolar⁵

Processo NFE (Numero de anos além da duração normal por curso)

Descrição do Processo: Este processo de negócio visa saber o número de anos para conclusão de um curso para além da normal duração de cada curso.

Granularidade: A granularidade usada na modelação deste processo de negócio é Fim Ano.

⁵ Estão contabilizados todos os anos curriculares que os estudantes frequentaram em currículos anteriores, por exemplo: estudantes que fizeram bacharelatos pré-bolonha estão contabilizados 3anos.

Tabela de Factos: A tabela de factos receberá uma linha e soma o total de alunos distribuídos por Ano Curricular, Curso e Ano Lectivo.

A tabela de factos Conta Alunos Ano Curricular contem informação sobre o número total de alunos por distribuídos por Ano Curricular, Curso e Ano Lectivo.

Dimensões: As dimensões que farão parte são: Cursos, Unidade Orgânica e Ano Lectivo.

	Anos												Total
	0	% 0	1	% 1	2	% 2	3	% 3	4	% 4	≥5	% ≥5	
ESTT	111	19,68%	49	8,69%	56	9,93%	61	10,82%	66	11,70%	221	39,18%	564
ESGT	70	30,97%	17	7,52%	51	22,57%	2	0,88%	42	18,58%	44	19,47%	226
ESTA	23	35,38%	6	9,23%	7	10,77%	2	3,08%	9	13,85%	18	27,69%	65
IPT Total	204	23,86%	72	8,42%	114	13,33%	65	7,60%	117	13,68%	283	33,10%	855

Tabela 19 - Número de anos para conclusão de curso

Por curso devemos saber quantos anos são necessários para a sua conclusão.

Número de Diplomados

Processo NDACE (Numero de diplomados por Ano lectivo/curso/Escola)

Descrição do Processo: Este processo de negócio visa saber o número de Diplomados por Ano Lectivo, Curso Escola.

Granularidade: A granularidade usada na modelação deste processo de negócio é o fim de cada ano.

Tabela de Factos: A tabela de factos receberá uma linha e soma o total de alunos diplomados por Ano Lectivo/Curso/Escola.

Dimensões: As dimensões que farão parte são: Alunos, Cursos, Unidade Orgânica e Ano Lectivo.

Correspondência no A3e's: 7.1.1 (Eficiência formativa).

Nº Diplomados	Ano Lectivo 1	Ano Lectivo 2	Ano Lectivo 3
ESTT	564	535	433
Curso A			
ESGT	226	237	233
Curso B			
ESTA	65	75	47
Curso C			
IPT Total			

Tabela 20 - Número de Diplomados

Número de Abandonos

Processo NAMACE (Numero de anulações de Matrícula por Ano lectivo/curso/Escola)

Descrição do Processo: Este processo de negócio visa saber o número de Anulações de Matricula por Ano Lectivo, Tipo de Aluno/Curso Escola.

Granularidade: A granularidade usada na modelação deste processo de negócio é Fim Ano.

Tabela de Factos: A tabela de factos receberá uma linha e soma o total de alunos que anulam a matricula por Ano Lectivo/Curso/Escola.

Dimensões: As dimensões que farão parte são: Alunos, Cursos, Unidade Orgânica e Ano Lectivo.

Total dos alunos que desistem de estudar;

	2005/2006			2006/2007			2007/2008		
	Total Alunos	Abandono	%	Total Estudantes	Abandono	%	Total Estudantes	Abandono	%
ESTT	1778	18	1,01%	1652	-29	-1,76%	1955	357	18,26%
ESGT	990	112	11,31%	998	91	9,12%	1159	19	1,64%
ESTA	415	49	11,81%	484	110	22,73%	459	104	22,66%
IPT Total	3183	179	5,62%	3134	172	5,49%	3573	480	13,43%

Tabela 21 Taxas de abandono

Rácio Aluno/Professor

Rácio Aluno/Professor por ano lectivo Escola/Curso

Processo RAP (Rácio de Alunos Professor)

Descrição do Processo: Este rácio visa saber o número de alunos por Professor em cada Escola/Curso por cada ano lectivo.

Granularidade: A granularidade usada na modelação deste processo de negócio é o fim de cada ano lectivo.

Tabela de Factos: A tabela de factos receberá uma linha e soma o total de alunos diplomados por Ano Lectivo/Curso/Escola.

Dimensões: As dimensões que farão parte são: Alunos, Cursos, Unidade Orgânica, Ano Lectivo.

	2005/2006	2006/2007	2007/2008	2008/2009
ESTT	12,75	12,71	16,20	15,71
ESGT	16,50	17,42	23,27	21,47
ESTA	8,79	12,57	10,98	10,40
IPT Total	12,90	13,88	16,97	16,13

Tabela 22 - Rácio Aluno/Professor

Serviços de Acção Social

Processo BEC (Bolsas de estudo concedidas por ano lectivo)

Descrição do Processo: Este processo de negócio visa saber o número de bolsas de estudo solicitadas, concedidas e número de estudantes em residência por ano lectivo em cada escola.

Granularidade: A granularidade usada na modelação deste processo de negócio é o fim de cada ano lectivo.

Tabela de Factos: A tabela de factos receberá uma linha e soma o total de alunos com bolsa de estudo concedida Ano Lectivo/ Escola.

Dimensões: As dimensões que fazem parte são: Alunos, Unidade Orgânica, Ano Lectivo.

	AnoLectivoN			AnoLectivoN+1			AnoLectivoN+2		
	S	C	R	S	C	R	S	C	R
ESTT									
ESGT									
ESTA									
IPT Total									

Tabela 23 - Evolução Bolsas de Estudo

(S) – Bolsas solicitadas, (C) – Bolsas Concedidas e (R) – Bolsas rejeitadas

Regime dos Maiores de 23 anos

Processo RM23 (Regime de Maiores de 23 Anos)

Descrição do Processo: Este processo de negócio visa saber o número de candidatos que tentam ingressar pelo regime de maiores de 23 anos.

Granularidade: A granularidade usada na modelação deste processo de negócio é o fim de cada ano lectivo.

Tabela de Factos: A tabela de factos receberá uma linha e soma o total de candidatos por tipo de ingresso maiores de 23 anos.

Dimensões: As dimensões que farão parte são: Candidato, Unidade Orgânica, Ano Lectivo.

	2006/2007	2007/2008	2008/2009
ESTT	69	154	90
ESGT	85	155	122
ESTA	64	43	41
IPT Total	218	352	253

Tabela 24 - Numero de Candidatos Regime M23

Quadro IV.36: Admissões pelo regime de M23 - 2008/2009

	Candidatos	Vagas	Matriculas	% Matriculas	% do Total de matriculas do 1º ano, 1ª vez
ESTT	72	75	58	80,56%	13,00%
ESGT	112	112	104	92,86%	26,80%
ESTA	41	30	29	70,73%	15,85%
IPT Total	153	142	133	86,93%	13,08%

Tabela 25 - Nº de Candidatos, Vagas por regime de ingresso M23

Prestação de Serviços à comunidade

Processo PSC (Prestação de Serviços à Comunidade)

Descrição do Processo: Este processo de negócio visa saber os valores das prestações de serviço à comunidade.

Granularidade: A granularidade usada na modelação deste processo de negócio é o fim de cada mês.

Tabela de Factos: A tabela de factos receberá uma linha com o valor das prestações de serviço por curso, unidade orgânica, ano.

Dimensões: As dimensões que farão parte são: Curso, Unidade Orgânica, Ano.

	Ano Lectivo 1	Ano Lectivo 2	Ano Lectivo 3
ESTT..ESGT..ESTA			
Curso A..Curso B...			
Prestação Serviços A			
IPT Total Prestação Serv.	€	€	€

Tabela 26 – Valor das prestações Serviço

5.4.2 DESENVOLVIMENTO TESTES E DOCUMENTAÇÃO

Depois de identificados os indicadores, efectuada a respectiva modelação dimensional, identificada e modelada a informação que é necessário extrair, damos início a esta actividade.

É nesta fase, que o repositório do *Data Warehouse* é construído e carregado com os registos informacionais operacionais através do processo de *ETL* (*Extraction, Transformation and Load*). Este processo é o responsável pela carga de dados e sua actualização a partir dos dados operacionais.

Para a extracção dos dados dos diversos sistemas fonte, optou-se por desenvolver um conjunto de scripts⁶ em linguagem *SQL* conforme atestam os anexos.

As principais operações nesta actividade de *ETL* são:

- Extracção - Processo que tem como função extrair os dados das fontes de dados (BD operacionais e fontes externas) e que deve causar o menor impacto nos Sistemas Operacionais;
- Transformação - Ocorre após a extracção e numa área à parte designada como área de estágio dos dados. Diz respeito à transformação a que os dados poderão estar sujeitos na passagem dos sistemas operacionais para o DW;
- Carregamento - Esta operação tem como actividade o carregamento no DW.

Existem, contudo, alguns problemas que passo a identificar e que tornam o ETL como uma tarefa árdua e complexa:

- No processo de extracção, os dados necessitam de ser limpos e convertidos. Datas, medidas e padrões precisam de estar normalizados. Exemplo: as datas na BD operacional de alunos SIGES pode estar no formato AA/MM/DD e noutra BD como o SIAG podem estar no formato MM/DD/AA. No *data warehouse* foi adoptado o padrão DD/MM/AAAA, então todos os dados necessitam de ser limpos para serem carregados no DW;
- É necessário eliminar a redundância de dados. Um mesmo aluno é inserido em aplicações diferentes. Ao entrar no DW, esse aluno só deverá aparecer uma única vez;
- Falta de documentação no dicionário de dados dos programas usados como fonte.

⁶ Os scripts são em termos práticos uma componente do programa que permite a realização de tarefas automaticamente.

6. CONCLUSÕES E RECOMENDAÇÕES

6.1 CONCLUSÕES

Este trabalho apresenta algumas contribuições importantes a todos os intervenientes no processo de desenho e construção de um armazém de dados. Entre estas contribuições, destacamos:

1 – Conforme definido no objectivo “Apresentar os benefícios da utilização de um DW como suporte à tomada de decisão”.

É de levada importância a criação dum repositório de dados, visto que um dos aspectos que gera mais confusão numa organização, é a informação muitas vezes contraditória que existe, assente em diversas aplicações suportadas em diferentes bases de dados. Verificou-se que muitas informações desejadas pelos utilizadores não estão disponíveis nos sistemas operacionais duma organização, levando em consideração que estes sistemas não foram criados para responder a necessidades analíticas. Ou, quando estão disponíveis, os dados estão dispersos por mais que uma aplicação, tornando difícil a sua recolha de uma forma correcta e agregada.

A criação dum DW pretende agregar todos os dados da organização num único repositório, de forma a disponibilizar as informações correctas aos utilizadores. A grande vantagem da implementação do DW é propiciar aos gestores uma significativa economia de tempo e esforço, no processo de tomada de decisão, comprovar indicadores de gestão da instituição e facilitar a disponibilização de dados, como por exemplo na disponibilização dos dados para o Guião do A3E's.

2 – Propor uma metodologia genérica para a implementação de um sistema de DW.

No Capítulo 3 foram apresentadas um conjunto de metodologias, e no capítulo 6 um desenho metodológico para a implementação dum Sistema de *Data Warehouse*, onde foram detalhadas as várias actividades e componentes desta metodologia.

3 - Identificar os processos de fornecimento de informações numa IES, listando os aspectos que contribuem para a elaboração do modelo de *Data Warehouse*.

No levantamento de requisitos foram identificados processos que podem necessitar de informação externa, como por exemplo processo relativo ao candidato em que informação dos candidatos nem sempre está no instituto, outro exemplo diz respeito percentagem de diplomados que obtiveram emprego em sectores de actividade relacionados com a área do ciclo de estudos, a informação sobre o emprego pode não existir no IPT. A dificuldade de muita das vezes poder aceder a esta informação é também um dos problemas identificados.

4 - Identificar os principais problemas que surgem com a extracção de dados dos actuais sistemas de informação, nas IES.

Um dos principais problemas com que nos deparamos é a qualidade da informação. Estamos-nos a referir à qualidade da informação operacional, pois deparamo-nos inúmeras vezes com situação em que os dados nos processos de extracção não estão totalmente carregados. As fontes operacionais muitas das vezes podem surpreender quer pela positiva (terem armazenada a informação esperada e até terem mais informação do que a esperada) quer pela negativa (haver falta de informação histórica).

Existe também a qualidade da informação no *Data Warehouse*, e aqui o problema surge quando se comparam dados do *Data Warehouse* com os dados transaccionais e o resultado não é coerente.

Relativamente às bases de dados analisadas, o facto de não existir documentação, como por exemplo, descrição de relações, diagramas de entidade relacionamento (DER), torna o seu entendimento demasiado confuso sendo desta forma um entrave à sua análise.

Foi identificado na ficha de levantamento de sistemas fonte que dados devem ser enviados para o *Data Warehouse*. Existem cuidados a ter na integração da informação, é esta actividade que permite construir o processo de ETL, permitindo assim a passagem da informação dos registos informacionais operacionais, para o *Data Warehouse*. A ênfase é garantir que a informação que está no *Data Warehouse* esteja devidamente integrada.

Independente da arquitectura adoptada, a maioria dos *Data Warehouses* é hoje construída sobre SGDB's relacionais. Vários problemas podem surgir na implementação de DW's quando se utilizam bases de dados relacionais, decorrentes das

diferenças existentes entre aplicações convencionais transaccionais *OLTP* e aplicações analíticas *OLAP*.

De forma a poder justificar a implementação de um Sistema de *Data Warehouse*, é importante começar por identificar a necessidade organizacional. Sendo a necessidade organizacional um factor fundamental para se justificar a necessidade de implementação dum Sistema de *Data Warehouse*. Deve também haver uma identificação clara e precisa das necessidades de negócio, pois sempre que se inicia um processo de DW este deve partir sempre da necessidade de negócio e nunca do lado das tecnologias de informação. É conveniente que o sistema de *Data Warehouse* tenha como objectivo a satisfação de uma necessidade (dados para o Guião de Avaliação de cursos A3E's), e que por cada necessidade haja um responsável (por exemplo o Director de Serviços ou de Unidade Departamental).

Devemos procurar o alinhamento do projecto com os objectivos organizacionais, de forma a permitir combinar várias orientações: objectivos, pedidos, processos, dados e tecnologia. Assim podemos colmatar deficiências existentes em cada uma das orientações, de modo a validar os vários requisitos informacionais, para satisfazer as necessidades organizacionais e dos seus colaboradores.

Este projecto teve como objectivo inicial, propor, construir e implementar um repositório de dados que contemple o fornecimento de informações precisas e consistentes no âmbito duma Instituição de Ensino Superior. Iniciou-se com a escolha duma necessidade organizacional, avaliação dos cursos ministrados no IPT, tendo como documento base o Guião de Avaliação disponibilizado pelo A3E's, que resultou num conjunto de *Data Mart's* da instituição, os *Data Mart's* de Avaliação. Outras áreas ou processos de negócio do politécnico certamente serão desenvolvidos e implementados, resultando no *Data Warehouse* final do IPT.

Com este projecto pretendeu-se atingir toda a instituição, apesar de podermos ter a impressão de estarmos restritos apenas a uma parte da organização, Gabinete de Avaliação e Qualidade, mas à medida que o trabalho se for desenvolvendo, irá sobressair a perspectiva de abrangência de toda a instituição.

É importante salientar que os processos de *data warehousing* estão em constante evolução e que apesar das dificuldades e problemas relatados, a evolução e disseminação das tecnologias de DW, tornam esses processos cada vez mais efectivos.

Os utilizadores, devem ser o mais envolvidos possível no processo de decisão de implementação do Sistema de *Data Warehouse*.

A criação de um *Data Warehouse* justifica-se por dois motivos fundamentais:

- A necessidade de integração de dados distribuídos por diferentes estruturas de bases de dados, tendo em vista uma análise global;
- A necessidade de separação dos dados utilizados nas operações correntes dos utilizados na análise e no reporting, tendo como fim a tomada de decisões.

Finalmente, é importante que nesta actividade se definam regras de acesso à informação, de forma a identificar quem pode aceder a que tipo de informação, e, para que fins essa informação é utilizada. Pois, nestes ambientes lida-se com informação sensível, e que pode ser utilizada para outros fins distintos daqueles a que estava previsto.

Data Warehousing é uma aplicação crítica por natureza. Os conjuntos de dados criados, utilizados e manipulados neste tipo de ambientes são da ordem de muitos gigabytes a alguns terabytes (Kimball R. , *The Data Warehouse Lifecycle Toolkit*, 1998).

Além disso, como o ambiente de Data Warehousing é utilizado por muitos utilizadores em simultâneo, obriga a um elevado grau de escalabilidade e paralelismo dos sistemas onde será alojado o DW. Desta forma neste tipo de ambiente os recursos de hardware e software são um dos principais factores de sucesso/fracasso (Kimball, L., M., e W., 1998).

6.2 TRABALHOS FUTUROS

Poderá dizer-se que este trabalho é de certa forma rico em oportunidades de desenvolvimento laterais para lá da implementação do *Data Warehouse*.

Fica, no entanto, como fulcral a oportunidade da implementação de novos *Data Mart's*, incorporação de outras fontes de dados, como os dados do Sistema Integrado de Apoio à Gestão, Sistema de Avaliação de Docentes ampliando o âmbito dos resultados. Podem também ser desenvolvidos novos relatórios, para que se possam extrair novas informações, como por exemplo, para a ficha de avaliação de docente, que foi identificada no levantamento de requisitos, mas não foi considerada como objecto de análise e modelação, e a recomendação da realização de um estudo mais aprofundado sobre tecnologias de *Data Warehouse* existentes no mercado, como trabalhos futuros.

REFERÊNCIAS BIBLIOGRÁFICAS

- Augusto Mateus e Associados. (2007). *Plano de Desenvolvimento do Instituto Politécnico de Tomar*. Tomar: IPT.
- Barquini, R. (1996). *Planning and designing the Warehouse*. New Jersey: Prentice-Hall.
- Caldeira, C. P. (2008). *Data Warehousing, Conceitos e Modelos*. Lisboa: Sílabo.
- Chadhuri, S. e. (1997). *An Overview of Data Warehousing and OLAP Technology*. New York: Sigmod Record.
- Chaves, M. C. (2010). *Pedagogia no Ensino Superior*. Coimbra: Formasau - Formação e Saúde, Lda.
- Conselho Coordenador Institutos Superiores Politécnicos. (2006). *Breve Caracterização do Ensino Superior em Portugal*. Lisboa: CCISP.
- Craig, R. S., Vivona, J. A., Hurwitz, J., & Bercovich, D. (1999). *Microsoft® Data Warehousing: Building Distributed Decision Support Systems*. Willey.
- Devlin, B. (1997). *Data Warehouse: from architecture to implementation*. Addison-Wesley.
- Filomena Castro Lopes, M. P. (2009). Desenvolvimento de Sistemas de Informação. In M. P. Filomena Castro Lopes, *Desenvolvimento de Sistemas de Informação* (p. 221). Lisboa: FCA.
- Gouveia, L. B. (2004). *Sistemas de Informação de Apoio à Gestão*. Porto: SPI.
- Harjinder, G., e RAO, P. C. (1996). *The Official Guide to Data Warehousing*. Que Corporation.
- Inmon, W. (2002). *Building the Data Warehouse*. John Wiley and Sons, Inc.
- Inmon, W. H. (1996). *Como construir o Data Warehouse*. São Paulo: Campus.
- Inmon, W. H. (1997). *Como Construir o Data Warehouse*. Rio de Janeiro: Campus.
- Inmon, W. H., Welch, J. D., & Glassey, K. L. (1999). *Gerenciando Data Warehouse*. São Paulo: Makron Books.
- Kimball, R. (1996). *Data Warehouse Toolkit*. São Paulo: Makron Books.
- Kimball, R. (1996). *The Data Warehouse Toolkit*. New York: John Wiley & Sons Inc.

- Kimball, R. (1998). *The Data Warehouse Lifecycle Toolkit*. Wiley.
- Kimball, R. (2002). *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling (Second Edition)*. Wiley.
- Kimball, R. (2004). *The Data Warehouse ETL Toolkit*. Wiley.
- Kimball, R., L., R., M., R., e W., T. (1998). *The Data Warehouse Lifecycle Toolkit: Expert Methods for Designing, Developing, and Deploying Data Warehouses*. John Wiley & Sons Inc.
- Le Moigne, J. L. (1974). *Les systèmes de decision dans les organisations*. Paris: Press universitaires de France.
- Macaulay, L. A. (1996). *Requirements engeneering*. London: 1996.
- Machado, F. N. (2004). *Tecnologia e Projecto de Data Warehouse*. São Paulo: Edições Érica.
- Reis, R. L., e Reis, H. P. (2008). *Gestão Estratégica, Aplicada a Instituições do Ensino Superior*. Lisboa: Escola Editora.
- Rosini, M. A., e Palmisano, A. (2003). *Administração de Sistemas de Informação e a Gestão do Conhecimento*.
- Santos, V. N. (2004). *Projecto e implementação de sistemas de Data Warehousing*. Minho, Portugal.
- Sezões, C., e Baptista, J. O. (2006). *Bussiness Intelligence*. Porto: SPI – Sociedade Portuguesa de Inovação.
- Soares, V. d. (1998). *Modelagem Incremental no ambiente de Data Warehouse*. Rio de Janeiro: UFRJ.
- Valente, D. L. (1996). *Estudo sobre Armazém de Dados*. Porto Alegre: CPGCC da UFRGS.
- Vieira, M. V. (2001). *Proposta de um Modelo de Sistema para Monitoramento da EAD em Cursos de Nível Superior Utilizando Ferramentas de Tecnologia da Informação*. UFSC.

ANEXOS

ANEXO A: GUIÃO DE AUTO AVALIAÇÃO DOS CICLOS DE ESTUDOS A3Es

ANEXO B: CARTA A SOLICITAR AUTORIZAÇÃO PARA A RECOLHA DE DADOS

GI - J. Pombo

CLASS.	01/2004/14
--------	------------

Joaquim Pombo da Silva Dias
Gabinete de Informática
Instituto Politécnico de Tomar
e-mail: jpombo@ipt.pt

05 06 03	3290/09
26 06 09	103/125/02

Ao Presidente do Instituto Politécnico de Tomar
Professor Dr. António Pires da Silva

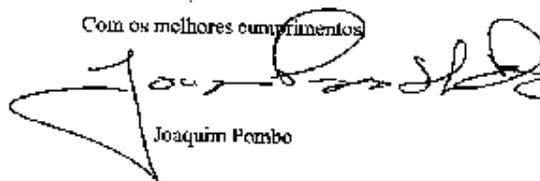
Assunto: Projecto de Investigação do Mestrado de Gestão - Sistemas de Informação no IPT

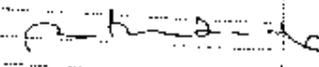
Sendo funcionário do Instituto Politécnico de Tomar, mais propriamente no Gabinete de Informática, encontro-me presentemente a realizar o meu projecto de Mestrado no Departamento de Gestão de Empresas da Universidade de Évora.

Uma das dificuldades que normalmente se verifica na extracção de informação numa organização seja ela de Ensino ou não deve-se à falta de integração dos dados existentes que se encontram em diversos formatos: ficheiros provenientes de processadores de texto, tabelas, folhas de cálculo, formulários, bases de dados, entre outras fontes.

No projecto de mestrado, que me proponho realizar, pretendo construir um modelo de armazém de dados (Data Warehouse) com o objectivo de organizar, armazenar e integrar as informações provenientes de outros formatos e sistemas, numa única base de dados para uma futura utilização no suporte à tomada de decisão no Instituto Politécnico de Tomar. Desta forma solicita-se autorização para a realização deste projecto utilizando como referência a instituição Instituto Politécnico de Tomar.

Com os melhores cumprimentos


Joaquim Pombo

DESPACHO

29.06.09

ANEXO C: ALGUNS COMANDOS *SQL* UTILIZADOS NA EXTRACÇÃO DE DADOS

ALUNO

```
SELECT
  CSE.T_ALUNOS.CD_ALUNO,
  SIGES.T_INDIVIDUO.NM_COMPLETO,
  SIGES.T_INDIVIDUO.DT_NASCIMENTO,
  SIGES.T_INDIVIDUO.SEXO,
  SIGES.T_INDIVIDUO.ESTADO_CIVIL,
  SIGES.T_INDIVIDUO.CD_NACIONA,
  SIGES.T_INDIVIDUO.CD_NACIONA_2,
  SIGES.T_INDIVIDUO.CD_NATURAL,
  CSE.T_HIST_INGRESSO.CD_INGRESSO,
  CSE.T_HIST_INGRESSO.CD_HAB_ANT,
  CSE.T_HIST_INGRESSO.CD_INST_HAB_ANT,
  CSE.T_HIST_INGRESSO.CD_TIP_EST_SEC,
  CASE
    WHEN CSE.T_TIPALUNO.CD_TIP_ALU = 4
      THEN 'Sim'
    ELSE 'Nao'
  END AS TRAB_ESTUDANTE,
  CASE
    WHEN CSE.T_HISTALUN.CD_SIT_BOLSA = 2
      THEN 'Sim'
    ELSE 'Nao'
  END AS BOLSA
FROM
  CSE.T_ALUNOS,
  SIGES.T_INDIVIDUO,
  CSE.T_HIST_INGRESSO,
  CSE.T_HISTALUN,
  CSE.T_TIPALUNO
WHERE
  CSE.T_ALUNOS.ID_INDIVIDUO = SIGES.T_INDIVIDUO.ID_INDIVIDUO AND
  CSE.T_ALUNOS.ID_ALUNO = CSE.T_HIST_INGRESSO.ID_ALUNO AND
  CSE.T_ALUNOS.CD_ALUNO = CSE.T_HISTALUN.CD_ALUNO AND
  CSE.T_HISTALUN.CD_LLECTIVO = CSE.T_TIPALUNO.CD_LLECTIVO AND
  CSE.T_HISTALUN.CD_CURSO = CSE.T_TIPALUNO.CD_CURSO AND
  CSE.T_HISTALUN.CD_ALUNO = CSE.T_TIPALUNO.CD_ALUNO
```

NOTA

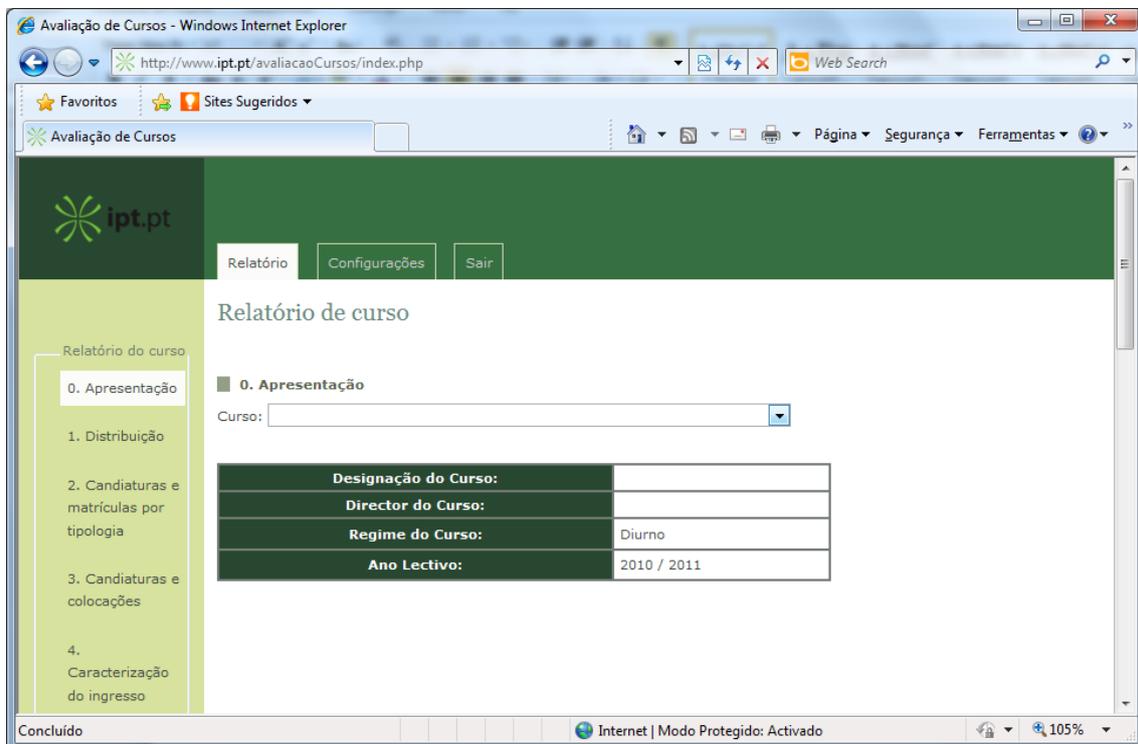
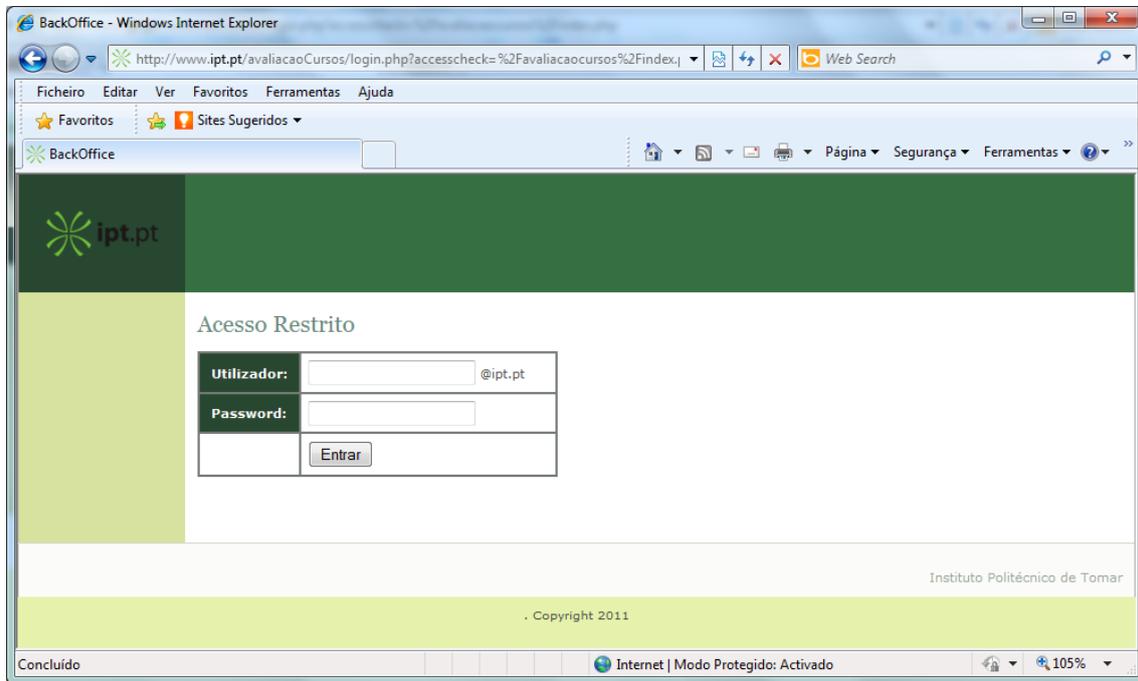
```
SELECT
  "CSE"."T_AVALUNO".CD_ALUNO,
  "CSE"."T_AVALUNO".CD_DISCIP,
  "CSE"."T_AVALUNO".CD_LLECTIVO,
  "CSE"."T_AVALUNO".CD_CURSO,
  "CSE"."T_AVALUNO".CD_DURACAO
FROM
  "CSE"."T_AVALUNO"
```

INSCRIÇÃO

```
SELECT
  "CSE"."T_INSCRI".CD_ALUNO,
  "CSE"."T_INSCRI".CD_DISCIP,
  "CSE"."T_INSCRI".CD_A_S_CUR,
  "CSE"."T_INSCRI".CD_STATUS
FROM
  "CSE"."T_INSCRI"
```


ANEXO D: EXEMPLOS DA INTERFACE DO PROGRAMA

Ecrã inicial da componente de *Front End*



Relatório de curso

0. Apresentação

Curso:

Designação do Curso:	<input type="text"/>
Director do Curso:	
Regime do Curso:	Diurno
Ano Lectivo:	2010 / 2011

Relatório de curso - 1

1. Distribuição de alunos inscritos no curso por anos (valor Abs. e %)

Anos lectivos	1º ano		2º ano		3º ano		Total	
	nº alunos	%	nº alunos	%	nº alunos	%	nº alunos	%
2007/2008	82	47.95	54	31.58	35	20.47	171	100
2008/2009	98	52.13	45	23.94	45	23.94	188	100
2009/2010	123	60.29	47	23.04	34	16.67	204	100
2010/2011	99	49.50	67	33.50	34	17.00	200	100

Avaliação de Cursos - Windows Internet Explorer

http://www.ipt.pt/avaliacaoCursos/index.php?page=relatorio§ion=candidatos

Avaliação de Cursos

ipt.pt

Relatório Configurações Sair

Relatório do curso

0. Apresentação

1. Distribuição

2. Candidaturas e matrículas por tipologia

3. Candidaturas e colocações

4. Caracterização do ingresso

2. Candidaturas e matrículas por tipologia de alunos

Anos Lectivos	Candidaturas								1º ano				1º ano 1ª vez		Cont. Geral	
	Cont. Geral		M23		Tit. CET		Outros		Total		1º ano		1º ano 1ª vez		Cont. Geral	
	nº alunos	%	nº alunos	%	nº alunos	%	nº alunos	%	nº alunos	%						
2007/2008	<input type="text"/>	0.00	<input type="text"/>	100	82	47.95	82	47.95	4	2.34						
2008/2009	<input type="text"/>	0.00	<input type="text"/>	100	98	52.13	98	52.13	17	9.04						
2009/2010	<input type="text"/>	0.00	<input type="text"/>	100	123	60.29	123	60.29	57	27.9						
2010/2011	<input type="text"/>	0.00	<input type="text"/>	100	99	49.50	58	29.00	9	4.50						

Concluído

Internet | Modo Protegido: Activado

Avaliação de Cursos - Windows Internet Explorer

http://www.ipt.pt/avaliacaoCursos/index.php?page=relatorio§ion=ingresso

Avaliação de Cursos

Relatório do curso

0. Apresentação

1. Distribuição

2. Candidaturas e matrículas por tipologia

3. Candidaturas e colocações

4. Caracterização do ingresso

5. Distribuição por género

6. Distribuição por faixa etária

7. Sucesso escolar

4. Caracterização do ingresso (dados relativos às diferentes fases de acesso)

Ano Lectivo	Fases	Nº de vagas	Nº de candidatos	Nº de candidatos (1ª opção)	Nº de colocados	Nº de colocados (1ª opção)	Classificação dos colocados (média)	Classificação do último colocado
2007/2008	1ª	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
	2ª	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
	3ª	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
	Total	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
2008/2009	1ª	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
	2ª	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
	3ª	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
	Total	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
2009/2010	1ª	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
	2ª	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

Concluído

Internet | Modo Protegido: Activado

Avaliação de Cursos - Windows Internet Explorer

http://www.ipt.pt/avaliacaoCursos/index.php?page=relatorio§ion=genero

Avaliação de Cursos

Relatório do curso

0. Apresentação

1. Distribuição

2. Candidaturas e matrículas por tipologia

3. Candidaturas e colocações

4. Caracterização do ingresso

5. Distribuição por género

6. Distribuição por faixa etária

7. Sucesso escolar

5. Distribuição do nº de alunos por género

Anos lectivos	Masculino		Feminino		Total	
	nº alunos	%	nº alunos	%	nº alunos	%
2007/2008	153	89.47	18	10.53	171	100
2008/2009	167	88.83	21	11.17	188	100
2009/2010	182	89.22	22	10.78	204	100
2010/2011	181	90.50	19	9.50	200	100

Internet | Modo Protegido: Activado

Avaliação de Cursos - Windows Internet Explorer

http://www.ipt.pt/avaliacaoCursos/index.php?page=relatorio§ion=faixa-etaria

Avaliação de Cursos

Relatório do curso

0. Apresentação

1. Distribuição

2. Candidaturas e matrículas por tipologia

3. Candidaturas e colocações

4. Caracterização do ingresso

5. Distribuição por género

6. Distribuição por faixa etária

7. Sucesso escolar

6. Distribuição do nº de alunos por faixa etária

Anos lectivos	< 20 anos		20 a 22 anos		23 a 30 anos		> 30 anos		Total	
	nº alunos	%	nº alunos	%	nº alunos	%	nº alunos	%	nº alunos	%
2007/2008	37	21.64	40	23.39	78	45.61	16	9.36	171	100
2008/2009	57	30.32	37	19.68	79	42.02	15	7.98	188	100
2009/2010	64	31.37	41	20.10	71	34.80	28	13.73	204	100
2010/2011	29	14.50	59	29.50	88	44.00	24	12.00	200	100

Concluído

Internet | Modo Protegido: Activado

The screenshot shows a web browser window titled 'Avaliação de Cursos - Windows Internet Explorer'. The address bar displays the URL: <http://www.ipt.pt/avaliacaoCursos/index.php?page=relatorio§ion=abandono>. The page content includes a navigation menu with 'Relatório', 'Configurações', and 'Sair'. The main heading is 'Relatório de curso - Engenharia Informática'. A sidebar on the left lists course report sections: 0. Apresentação, 1. Distribuição, 2. Candidaturas e matrículas por tipologia, 3. Candidaturas e colocações, 4. Caracterização do ingresso, and 5. Distribuição. The main content area is titled '8. Taxa de abandono' and contains a table with the following data:

Anos lectivos	Total de alunos inscritos no curso	Abandono (nº alunos)	Abandono (%)
2007/2008	171	110	64.33
2008/2009	188	120	63.83
2009/2010	204	96	47.06
2010/2011	200	-52	-26.00

Below the table, the formula for calculating the abandonment rate is provided:
$$\text{Abandono Ano (n)} = \text{Total de alunos inscritos no Ano (n)} - (\text{Total alunos inscritos Ano(n-1)} + \text{nº novos alunos Ano (n)} - \text{nº diplomados Ano (n)})$$

The browser's status bar at the bottom indicates 'Internet | Modo Protegido: Activado' and a zoom level of 105%.

ANEXO E: EXEMPLOS DE MODELAÇÃO UTILIZANDO O VISIO

