
MGI

MESTRADO
Gestão de Informação

***PROPOSTA DE UM MODELO CONCEPTUAL DE
DATA WAREHOUSE PARA O INSTITUTO
NACIONAL DE ESTATÍSTICA DE MOÇAMBIQUE***

Calado Pereira Fijamo

Dissertação apresentada como requisito parcial para
obtenção do grau de Mestre em Gestão de Informação

Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

**PROPOSTA DE UM MODELO CONCEPTUAL DE DATA
WAREHOUSE PARA O INSTITUTO NACIONAL DE ESTATÍSTICA DE
MOÇAMBIQUE**

por

Calado Pereira Fijamo

Dissertação apresentada como requisito parcial para a obtenção do grau de Mestre em Gestão de Informação, Especialização em gestão de sistemas e tecnologias de informação

Orientador: Professor Doutor Miguel de Castro Neto

Novembro, 2012

DEDICATÓRIA

A Deus pela iluminação e orientação quotidiana

À minha avó Madalena Fijamo que infelizmente não se encontra entre nós, pelos ensinamentos e paciência em me criar.

Ao meu tio Oliveira Fijamo que foi um autêntico pai.

À minha esposa Leta Alberto Matola, pela compreensão e paciência ao longo destes anos.

Aos meus filhos Maria da Gloria, Ilton e Calado Júnior, pelo entendimento, paciência e privações que passaram ao longo destes anos.

AGRADECIMENTOS

Agradeço em primeiro lugar Deus todo-poderoso, por tudo o que proporcionou até esta data, e o que proporcionará no futuro.

Agradeço ao meu orientador, professor Doutor Miguel de Castro Neto, pelos conhecimentos transmitidos neste trabalho e durante toda minha estadia como estudante. Agradecer igualmente pela sua forma especial e incomparável de lidar com seus estudantes.

Ao Instituto Nacional de Estatística, em particular ao Dr. João Dias Loureiro (Presidente do INE) pela confiança e oportunidade depositada.

Agradeço ao professor Doutor Pedro Cabral, por todo o apoio que proporcionou durante a minha carreira como estudante.

Agradeço igualmente a todos professores e funcionários do ISEGI, pelos ensinamentos e apoio prestado.

Ao meu amigo João Gabriel Loforte, pelo apoio moral e social.

Aos meus colegas da turma, em especial ao Nuno Dionísio e Daniel Heitor.

A todos os que direta ou indiretamente contribuíram para o desenvolvimento deste trabalho.

RESUMO

O mercado atual é bastante competitivo, com constantes exigências em transformação de dados em informação e esta por sua vez em conhecimento.

Uma informação clara e válida é o grande segredo de qualquer ramo de atividade, e é por isso que o conhecimento é a chave fundamental na tomada de decisão.

O *Data Warehouse* desempenha um papel primordial como ferramenta de apoio na tomada de decisão, apresentando avanços substanciais comparativamente aos sistemas ou bases de dados tradicionais, oferecendo maior fiabilidade aos decisores estratégicos na sua difícil missão de tomar decisões rápidas e seguras.

O instituto nacional de estatística de Moçambique, abreviadamente designado por INE-M, é o responsável pelas estatísticas oficiais do País, e detém um grande volume de dados de interesse nacional e internacional para produção de informação estatística. Estes dados, encontram-se armazenados e organizados em plataformas tradicionais, dificultando o acesso rápido e eficiente aos decisores quando pretendem tomar alguma decisão com segurança.

Neste contexto, torna-se importante a existência de um repositório seguro e de acesso rápido, com capacidade de armazenar dados históricos da instituição.

Este repositório facilitará aos dirigentes o cumprimento das suas obrigações e deveres de tomada de decisão, de forma eficaz e consistente.

Um modelo conceptual de *Data Warehouse* que sirva de apoio estratégico na tomada de decisão será de certa forma a melhor e mais conceituada contribuição que o INE-M pode oferecer aos seus utilizadores e parceiros, visto que, terão sempre resposta segura e em tempo útil.

PALAVRAS-CHAVE

Modelo conceptual de *Data Warehouse*, *Data Warehouse*, *Data Marts*, Base de dados.

ABSTRACT

The current markets are very competitive, with constant demands for transforming data into information and turn it into knowledge.

Clear and valid information is the great secret of any industry, which is why knowledge is the key decision-making factor.

The data warehouse plays a key role as a support tool in decision making, representing a substantial progress, when compared to systems or databases considered traditional, giving more reliability to the strategic decision-makers in their difficult task of making fast and accurate decisions.

The national statistical institute of Mozambique, hereinafter called INE-M, manages the official statistics of the country, and holds a large volume of data of national and international interest for the production of statistical information. This data is stored and organized in traditional platforms, making it's access difficult quick and efficient decision-makers when you want to make consistent decisions.

In this context, it is important the existence of a repository for secure and fast access, capable of storing historical data of the institution.

This repository makes it easier for leaders to fulfill their obligations and duties of effective and consistent decision-makinging.

A conceptual model of data warehouse that serves to support strategic decision making, is somehow the best and most respected contribution that the INE-M can offer to their users and partners, since they will always respond in a timely and secure manner.

KEYWORDS

Conceptual model Data Warehouse, Data Warehouse, Data Marts, Database.

ÍNDICE

1. Introdução.....	1
1.1. Definição do Tema	2
1.2. Organização do Trabalho	2
1.3. Objetivos do Trabalho.....	3
1.3.1. Objetivo geral.....	3
1.3.2. Objetivos específicos	3
1.3.3. Interesse na investigação.....	3
1.4. Fases de Elaboração do Trabalho	4
2. Revisão da Literatura	5
2.1. Enquadramento Teórico	5
2.1.1. Dado	6
2.1.2. Informação	6
2.1.3. Conhecimento.....	7
2.1.4. Sistemas da informação e a evolução dos sistemas empresariais	7
2.2. Informação nas organizações	10
2.3. Sistemas Operacionais (OLTP) e Sistemas de Apoio à Decisão (SAD).....	13
2.4. Caracterização de Data Warehouse.....	15
2.5. Modelos e tabelas de Data Warehouse.....	20
2.6. Modelos de DW.....	21
2.6.1. Modelo em Estrela	21
2.6.2. Modelo em Floco de Neve	22
2.6.3. Modelo em Constelação	23
2.7. Processo de Modelação	24
2.8. Erros mais frequentes na modelação.	25
2.9. Ferramentas de Busca, Transformação e Acesso	27
2.9.1. ETL	28
2.9.2. Ferramenta OLAP	30
2.10. Business Intelligence (BI) e Data Mining.....	32
2.11. Custos e Benefícios do DW	34
2.11.1. Custos	34
2.11.2. Benefícios	35

2.12. Fatores críticos de Sucesso do DW	36
3. metodologia	38
3.1. Metodologia de Pesquisa.....	38
3.2. Metodologia conceptual de DW	41
3.2.1. Metodologia de INMON.....	42
3.2.2. Metodologia de Kimball.....	43
3.2.3. Comparação das Metodologias do INMON e KIMBALL.....	45
3.3. Requisitos do Data Warehouse.....	46
4. Instituto Nacional de Estatística de Moçambique	47
4.1. O INE – M como Organização.	47
4.2. Missão do INE – M	49
4.3. Visão do INE – M	49
4.4. Valores do INE – M.....	49
4.5. Desafios do INE – M	49
4.6. Necessidade de Informação.....	51
4.7. Gestão Estratégica	52
4.8. Requisitos do projeto.....	53
5. Modelação do Data Warehouse para o INE-M.....	56
5.1. Modelo Proposto	56
5.2. Modelo E-R de sistemas operacionais	58
5.3. Modelo dimensional de DW	61
6. Data Mart	64
6.1. Caraterização	64
6.2. Criação de DM.....	64
6.2.1. Seleção do processo de Negócio	65
6.2.2. Definição da Granularidade	65
6.2.3. Escolha das Dimensões	66
6.2.4. Identificação de factos.....	67
6.3. Modelação	67
6.3.1. Modelo conceptual	67
6.3.2. Modelo lógico	68
6.3.3. Projeto físico	70
7. Discussão de resultados.....	72
7.1. Projeto do Sistema de DW no INE-M.....	72

7.2. Participação, Localização e carga.....	72
7.3. Atualização e manutenção do DW.....	73
7.4. Consideração final.....	73
8. Conclusões	74
9. Limitações e Recomendações para Trabalhos Futuros	75
9.1. Limitações no estudo	75
9.2. Recomendações para trabalhos futuros.....	75
10. Bibliografia	76
11. Anexos.....	79

ÍNDICE DE FIGURAS

Figura 2.1.3 - Etapas do processo de conhecimento.....	7
Figura 2.2.1 - Tipos de SI.....	11
Figura 2.2.2 - Características da decisão.....	12
Figura 2.4.1 - Processo de data warehouse.....	17
Figura 2.4.2 - Arquitetura de Data Warehouse	17
Figura 2.4.3 - Elementos básicos de DW	19
Figura 2.5 - Dimensões e de facto com os seus atributos	21
Figura 2.6.1 - Modelo estrela.....	22
Figura 2.6.2 - Modelo Floco de Neve.....	23
Figura 2.6.3 - Modelo em constelação	23
Figura 2.9.1 - Processo ETL	29
Figura 2.9.2 – Carregamento e transformação de dados.....	30
Figura 3.2.2.2 – Arquitetura bottom-up	44
Figura 4.1.1 - Organograma do INE – M	48
Figura 4.1.2 - Estrutura organizacional do INE	49
Figura 4.7.1 - Perspetivas estratégicos	53
Figura 5.1.1 - DW para INE-M.	57
Figura. 5.1.2 - Modelo dimensional do DW em estrela para o INE-M	58
Figura 5.2.1 - Atributos e caracterização de cada entidade	59
Figura 5.2.2 - E-R, cardinalidade e hierarquia da produção Industrial.....	59
Figura 5.2.3 - E-R, cardinalidade e hierarquia de comércio e serviços.....	60
Figura 5.2.4 – Diagrama E – R Comércio e Serviços.....	60
Figura 5.2.5 – Diagrama E – R de estatísticas de Produção Industrial.....	61
Figura 5.2.6 - Exemplo de transformação e aglomeração de dados.	61
Figura 5.3.1 Categorias de entidades	62
Figura 5.3.2 - Hierarquia de um modelo conceptual.....	63
Figura 6.1.1 - Constituição de DM com Tabelas de facto e dimensionais.....	64
Figura 6.3.1.1 - Modelo conceitual das Estatísticas Empresariais.....	68
Figura 6.3.1.2 - Modelo conceitual com os atributos.....	68
Figura 6.3.2.1 - Modelo lógico das estatísticas Económicas e Empresariais.....	69
Figura 6.3.2.2 - Esquema do modelo lógico em Access.....	69
Figura 6.3.2.3 - Esquema do modelo lógico em SQL.	70

ÍNDICE DE TABELAS

Tabela 2.1.1 - Níveis de Informação	5
Tabela 2.1.2 – Caracterização da Qualidade de Informação	6
Tabela 2.1.4 - Evolução histórica	9
Tabela 2.3.1 - Limitação de dados OLTP	15
Tabela 2.7 - Aspetos no Processo de Modelação	25
Tabela 2.9.1 - Ferramentas de DW	28
Tabela 2.9.2 - Ferramentas OLTP e OLAP	32
Tabela 3.1.1 - Amostra de recolha de dados.....	40
Tabela 4.7.1 - Análise SWOT	52

ÍNDICE DE ANEXOS

Anexo 11.1 - Ficha de dados área de Comércio e Serviços.	79
Anexo 11.2 - Ficha de dados de Área industrial	80
Anexo 11.3 - Guião de apoio na recolha de requisitos (gestão de topo)	81
Anexo 11.4 - Guião de apoio na recolha de requisitos (gestão intermédia).....	82
Anexo 11.5 - Guião de apoio na recolha de requisitos (Departamento de Informática)	83

LISTA DE SIGLAS E ABREVIATURAS

BD	Base de Dados
BDMD	Base de Dados Multidimensionais
DARH	Departamento de Administração e Recursos Humanos
DB	<i>Database</i>
DCI	Departamento de Censos e Inquéritos
DCNIG	Departamento de Contas Nacionais e Indicadores Globais
DEBA	Departamento de Estatísticas de Bens e Agrícolas
DEMOVIS	Departamento de Estatísticas Vitais e Sociais
DESC	Departamento de Estatísticas Sectórias e Comércio
DIPC	Departamento de Índice de Preço e Conjuntura
DISI	Departamento de Informática e Sistemas de Informação
DM	<i>Data Mart</i>
DMs	<i>Data Marts</i>
dmg	<i>data mining</i>
DSS	Decision Support Systems
DW	<i>Data Warehouse</i>
EDW	<i>Enterprise Data Warehouse</i>
E-R	Entidade Relacionamento
ERP	Enterprise Resource Planning
ETC	Extração transformação e Carga
ETL	<i>Extract Transform Load</i>
GC	Gestão de conhecimento
IBGE	Instituto Brasileiro de Geografia e Estatística
INE-M	Instituto Nacional de Estatística de Moçambique
INE-PT	Instituto Nacional de Estatística de Portugal
IT	<i>Information Technology</i>
KM	<i>Knowledge Management</i>
MD	Meta dados

MER	Modelagem Entidade-Relacionamento
MIS	<i>Management Information Systems</i>
MOLAP	<i>Multidimensional OLAP</i>
MRP	Material Requirements Planning
MRP II	Manufacturing Resources Planning
OLAP	<i>Online Analytical Processing</i>
OLTP	<i>Online transaction Processing</i>
PARPA	Plano de Ação de Redução da Pobreza Absoluta
ROLAP	<i>Relational OLAP</i>
SAD	Sistemas de Apoio à Decisão
SEN	Sistema Estatístico Nacional
SGBD	Sistema de Gestão de Bases de Dados
SI	Sistemas de informação
SIE	Sistemas de Informação e Estatísticas
SO	Sistemas Operacionais
SSD	Sistemas de Suporte a Decisão
TI	Tecnologias de Informação
TIC	Tecnologias de Informação e Comunicação

1. INTRODUÇÃO

Este estudo tem como finalidade a dissertação de mestrado, e terá um papel preponderante para propor a criação de um repositório de dados que apoie a tomada de decisão no Instituto Nacional de Estatística de Moçambique (INE – M).

Atualmente as empresas ou instituições credíveis apostam em metodologias de gestão empresarial de dados. Consequentemente depara-se com um grande crescimento do volume de dados. Para atender a esse crescimento, é necessário que a empresa ou organização crie um ambiente informatizado sólido, com boa capacidade de armazenamento e gestão de dados.

O sistema de gestão de dados empresarial deve garantir a segurança, coesão, interatividade e acima de tudo, deve apoiar com eficiência os decisores na sua complexa missão de tomada de decisão.

Sendo o Instituto Nacional de Estatística de Moçambique, o responsável pela produção e coordenação das estatísticas oficiais do País, é este que responde às necessidades estatísticas dos utilizadores. Por isso, é detentor de grandes volumes de dados.

Nesta fase da globalização, para uma organização poder servir melhor os seus clientes, deve contar como aliados indispensáveis os sistemas e tecnologias de informação. Só assim, poderá enfrentar (ou promover, no caso do INE – M) a competitividade no mercado modernizado.

“As organizações que não introduziram na sua cultura conceitos de excelência nos seus produtos e serviços, incluindo o atendimento ao cliente, podem estar a comprometer a sua sobrevivência” (Santos, 2008).

O INE – M na qualidade de promotor da utilização da informação estatística oficial no País, deve possuir uma base de dados de alta qualidade, que garanta confiança e segurança, de forma a nunca comprometer a eficiência e imagem da instituição assim como os que dela dependem.

1.1. DEFINIÇÃO DO TEMA

Proposta de um modelo conceptual de *Data Warehouse* (DW). Este é o tema avançado como ferramenta de apoio na tomada de decisão para o INE-M. Pretende-se com o DW integrar todas fontes e suportes de informação dispersas pelos pelouros, num único repositório que proporcione o fácil e rápida resposta a questões de decisão.

1.2. ORGANIZAÇÃO DO TRABALHO

A organização deste trabalho obedece na primeira parte à seguinte ordem: dedicatória, agradecimentos, resumo, índice geral, índice de figuras, índice de tabelas, índice de anexos, lista de siglas e abreviaturas.

A segunda parte é composta por uma estrutura de 9 capítulos numerados sequencialmente de 1 a 9.

O primeiro capítulo compreende a introdução, definição do tema, organização do trabalho, objetivos e as fases de elaboração da dissertação.

O segundo capítulo é constituído pela revisão de literatura, que conta com o enquadramento teórico do tema, informação nas organizações, comparação dos sistemas tradicionais e de suporte a decisão, caracterização de um *data warehouse*, modelos de *data warehouse*, tabelas de (facto e dimensão), processos de modelação, erros frequentes na modelação, principais ferramentas de *data warehouse* (DW), *business intelligence* (BI) e *data mining* (DMg), custos e benefícios, e fatores críticos de sucesso.

O terceiro capítulo apresenta a metodologia do trabalho, que consiste na metodologia da pesquisa, metodologias de DW, descrição das metodologias de Inmon e de Kimball, comparação das diferentes abordagens destes dois autores e requisitos de DW.

O quarto capítulo aborda a instituição do estudo de caso, O INE-M, como organização, a sua visão, missão, desafios, necessidades de informação e fatores de sucesso.

O quinto capítulo contém a definição teórica do modelo conceptual de DW, apresenta o conceito do modelo dimensional e descreve o modelo proposto para o INE-M.

O sexto capítulo é a conceção dimensional.

O sétimo capítulo é a discussão de resultados

O oitavo capítulo é de conclusões.

O Nono capítulo é de limitações e recomendações.

1.3. OBJETIVOS DO TRABALHO

1.3.1. Objetivo geral

Propor ao INE – M um modelo conceptual de *Data Warehouse*, que sirva de apoio na tomada de decisão rápida e segura. Este objetivo surge como forma de resolver ou minimizar o problema que o Instituto Nacional de Estatística de Moçambique, vem enfrentando para responder atempadamente aos pedidos dos seus utilizadores.

1.3.2. Objetivos específicos

Este trabalho tem como objetivos específicos:

Propor um modelo que apoie os utilizadores da informação estatística em Moçambique.

Aumentar a capacidade de resposta dos técnicos do INE-M, na gestão dos dados estatísticos.

1.3.3. Interesse na investigação

Elevar o nível académico e competências de investigação.

Apoiar o INE – M com uma metodologia adequada para o apoio na tomada de decisão.

Atualmente as organizações governamentais ou não governamentais recorrem aos sistemas de informação, como forma de enfrentar o mercado competitivo. Assim, minimizam os possíveis erros da informação disponibilizada para o público.

Numa instituição quanto mais dados e informação existirem, mais dificuldades surgem no manuseio da mesma, piorando quando esses dados estiverem completamente dispersos na organização.

Normalmente estas organizações recebem dados vindos de diversos locais ou fontes externos, com formatações diversificadas e por vezes incompatíveis com a tecnologia existente na instituição.

Uma das alternativas de integração e uso desses dados é o DW, que é um repositório de dados fortemente normalizado.

Sendo o INE – M uma instituição do aparelho de estado com muitos dados da natureza acima referida, aconselha-se a existência de um DW capaz de extrair, integrar, tratar e guardar dados ou informação de forma segura e que possibilite um posterior uso no apoio às tomadas de decisão.

1.4. FASES DE ELABORAÇÃO DO TRABALHO

Fase 1 - Pesquisa bibliográfica / revisão da literatura sobre o tema, bem como análise de modelos conceptuais.

Fase 2 - Análise e especificação de requisitos funcionais do DW. É a fase de conhecimento da informação dos utilizadores finais. Será em função destes requisitos, e da análise dos sistemas operacionais existentes na empresa, que servirão de base à conceção do modelo de DW.

Fase 3 - Metodologia do trabalho, consiste na definição dos métodos utilizados na pesquisa, e a metodologia a usar na construção deste projeto de DW. Ainda nesta fase vai-se descrever as teorias e práticas das metodologias defendidas quer por Inmon quer por Kimball, prosseguindo com a comparação destas diferentes abordagens. Esta fase termina com a descrição dos requisitos básicos de DW.

Fase 4 – Descreve a instituição de estudo de caso, descreve igualmente a infraestrutura tecnológica existente na instituição, assim como, os requisitos básicos para implementar o sistema multidimensional a propor. É nesta fase, que se procede à análise das fontes de dados que irão alimentar o DW, tanto ao nível dos sistemas que os originam como dos próprios dados em si. Esta análise deve ser feita tendo em conta a facilidade de transformação e carregamento dos dados para o repositório a propor.

Fase 5 – Modelação de DW para o INE – M, é a definição do modelo conceptual e escolha do modelo dimensional para o estudo de caso.

Fase 6 – Data Mart, é a conceção de um dos DM do modelo de DW proposto.

Fase 7 – Discussão de resultados.

Fase 8 – Conclusões.

Fase 9 – Limitações e Recomendações.

2. REVISÃO DA LITERATURA

Esta revisão da literatura tem como propósito fundamental encontrar material que sirva de base à sustentação teórica da pesquisa para realização do trabalho de dissertação de mestrado.

Fazer um enquadramento teórico, histórico, e caracterizar o DW, com base no levantamento bibliográfico, com maior incidência em Inmon e Kimball, por serem os pioneiros neste tipo de modelo de armazenamento de dados.

Data Warehouse como grande repositório de dados, é um instrumento muito importante e indispensável nos negócios da empresa, num mercado moderno e competitivo onde os gestores devem ser ágeis e seguros na tomada de decisão.

2.1. ENQUADRAMENTO TEÓRICO

Para um melhor enquadramento do que é na essência o *data warehouse*, começa-se por definir genericamente dados, informação, conhecimento e Sistemas de Informação (SI). E em seguida, falar da evolução histórica dos SI.

Para se cumprir com êxito os objetivos preconizados numa instituição, é necessária a existência de uma informação muito bem organizada, com especial cuidado na qualidade da mesma. Qualquer informação tem o seu valor e importância consoante o nível de empregabilidade, seja operacional, tático e estratégico. Cada nível tem as suas necessidades de informação, para poder cumprir as diretrizes planificadas na organização. A tabela 2.1.1 mostra os níveis básicos da informação.

Nível	Descrição
Dados	Elementos atômicos que qualificam, referenciam e descrevem todos os passos necessários à operação do sistema de informação.
Informação	Dados agregados por meios de relações de complementaridade entre eles.
Conhecimento	É a forma ou mecanismo de como os dados e informação são transmitidos de maneira organizada e estruturada para o uso futuro.

Tabela 2.1.1- Níveis de Informação; Fonte: Adaptado de Gouveia (2004)

2.1.1. Dado

Considera-se como dado qualquer elemento ou símbolo que por si só não possui nenhum significado. “Factos que descrevem os objetos de informação (eventos e entidades). Os dados referem-se a mais de um facto. Um determinado facto é referido como item” (Gouveia, L. M. B., 1996, p.17).

2.1.2. Informação

É um conjunto de dados processados e com um significado ou sentido.

As palavras informação e dados, são ambas intercambiáveis em muitos contextos. Todavia, não são sinônimos. Por exemplo, de acordo com a observação de Adam M. Gadowski (1993), dados é tudo que pode ser processado e as informações são dados que descrevem um domínio físico ou abstrato. Knuth aponta que o termo dado se refere à representação do valor ou quantidade medida ao passo que informação, quando usada em um sentido técnico, é o significado daquele dado (Wikipédia, a enciclopédia livre, (consultada em 02/04/2012))

“Informação é um conjunto específico de dados com características próprias, isto é, conjunto de dados significativos e relevantes para o sistema a quem se destinam” (Gouveia, L. M. B., 1996, p.17).

Com as definições acima fica claro que, para ter informação é preciso necessariamente que se tenha dados. É com bons dados que se pode ter boa informação, por isso, a proveniência dos mesmos deve ser de confiança. A tabela 2.1.2 ilustra características de qualidade da informação obtida através de dados.

Característica	Qualidade da Informação
Precisa	Informação com grau de rigor que revela uma caracterização real, fiável, possível, correta e verdadeira.
Oportuna	A presença do fluxo de informação apropriado, contribui na velocidade de reação da organização.
Completa	A informação na organização tem valor quando disponível, e não quando esta dispersa.
Concisa	Informação pormenorizada ou muito extensa, não é utilizada por contrariar dois princípios básicos de comunicação: fácil de descodificar e fácil de difundir.

Tabela 2.1.2 – Caracterização da Qualidade de Informação; Fonte: Adaptado de Gouveia (2004)

2.1.3. Conhecimento

O conhecimento é complexo, cada autor tem perspectiva diferente sobre o assunto. O conhecimento pode ser encontrado em dois níveis, tácito e explícito.

Tácito – conhecimento pessoal, que não é facilmente formulado e expresso.

Explícito – conhecimento formalizado e codificado.

“Conhecimento é um processo, composto de várias etapas, não trivial, interativo e iterativo, para identificação de padrões compreensíveis, válidos, novos e potencialmente úteis a partir de grandes conjuntos de dados” (Fayyad et al, 1996). Este autor faz entender que o conhecimento útil deve ser extraído e derivado de padrões de dados, inter-relacionados por tarefas funcionais, com a finalidade de auxiliar nas tomadas de decisão.

“O interativo diz respeito ao controle humano sobre um processo de conhecimento no sentido de analisar e interpretar os dados e padrões, e o iterativo significa a possibilidade de repetições integrais ou parciais do processo, a fim de aperfeiçoar e alcançar os resultados satisfatórios” Carvalho, I. M.; Muñoz, D. L.C.; Ferreira, M. V. A. S.; Santos, P. M.; Rover, A. J.; Fialho, F. A. P. (ND). A figura 2.1.3 mostra fases do processo do conhecimento.

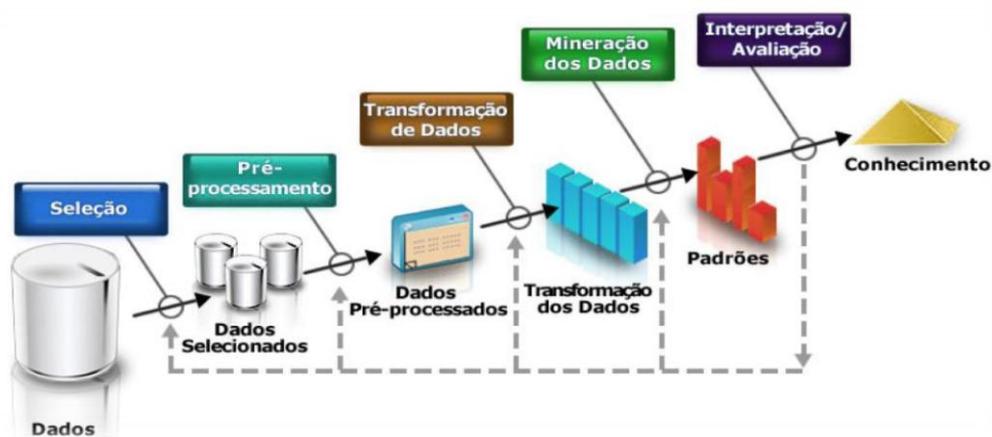


Figura 2.1.3 - Etapas do processo de conhecimento. Fonte: Adaptado de Fayyad (1996).

2.1.4. Sistemas da informação e a evolução dos sistemas empresariais

Sistema de informação (SI) é o elemento chave para o desenvolvimento de uma organização no mundo do mercado competitivo e inteligente. “A teoria de sistemas de

informação é a base de trabalho para os analistas de sistemas e um bom suporte para a compreensão duma área de negócio, para auditores, consultores e os próprios agentes decisores” (Gouveia, L. M. B., 1996, p.10).

Sistema de informação é um “conjunto de aplicações que partilham dados entre si” (Silva, M. M. da, 2003, p.2).

Pode-se observar que atualmente os sistemas de informação desempenham um papel de relevo no mundo empresarial, em outras palavras, uma empresa não poderá adquirir vantagens competitivas se não usar como suporte os sistemas de informação.

Com a modernização os sistemas passaram a estar diretamente relacionados com a vida da organização, quer no controlo quer nas atividades. A necessidade das organizações de possuírem dados integrados, históricos e ágeis, contribuiu positivamente para implementação e uso de DW. Ao longo do tempo, os sistemas foram evoluindo através das crescentes necessidades dos utilizadores, isso fez com que existisse a preocupação da busca de tecnologias cada vez mais avançadas para o alcance dos objetivos da organização.

É por isso que, considera-se tecnologia da informação a base de recurso estratégico para o desenvolvimento contínuo de vantagem competitiva.

Com a base das definições acima, pode-se fazer um breve historial da evolução dos SI empresarial ao longo das últimas décadas.

Os sistemas empresariais desenvolveram na mesma conformidade do desenvolvimento do *hardware* e do *software*. “Fazendo uma perspetiva de SI pode-se considerar em desenvolvimento bastante acelerado” (Hossain, L.; Patrick, J. D. & Rashid, M. A., 2002).

Na década 60, muitos sistemas informáticos centralizados foram desenvolvidos e implementados pelas organizações, com o intuito de efetuar o controlo no armazenamento de mercadorias, inventários e para cálculos contabilísticos da empresa. Nesta década, guardavam-se grandes volumes de mercadorias nos armazéns como forma de colmatar a demora na entrega de mercadoria ao cliente.

Nos anos 70, percebem que manter grandes volumes de mercadorias nos armazéns implicava maiores custos de manutenção dessas mercadorias, bem como a imobilização de uma quantidade importante de capital. Como forma de minimizar custos, houve a necessidade de repensar a gestão do armazenamento de matéria-prima e do produto acabado. Neste contexto, os programadores e consultores começaram a pesquisar uma melhor forma de rentabilizar o seu trabalho. “As empresas começaram a desenvolver seus próprios pacotes, que posteriormente foram adaptados à realidade de outras empresas, tendo culminado com o aparecimento dos *Material Requirements Planning Systems* (MRP). Estes sistemas permitiam gerir o produto ou os seus componentes de acordo com o programa da produção” (Hossain, L.; Patrick, J. D. & Rashid, M. A., 2002).

Esta geração de *software* chegou até aos meados dos anos 80, sucedido pelo *Manufacturing Resources Planning* (MRP II), que no seu funcionamento era muito semelhante ao seu predecessor, com a diferença de colocar maior ênfase no controlo de inventários, repercutindo-se diretamente na contabilidade e finanças.

Nos anos 90 deu-se o aparecimento dos sistemas ERP, que forneciam informação mais precisa que os MRP e MRPII. No sistema ERP as atividades do negócio da empresa eram todas concentradas no mesmo pacote, sendo registadas todas as transações efetuadas, enquanto que nas anteriores cada pacote poderia significar uma tarefa ou atividade. A tabela 2.1.4 mostra a evolução histórica dos sistemas empresariais.



2000s	<i>Extended ERP (XER)</i>
1990s	<i>Enterprise Resource Planning (ERP)</i>
1980s	<i>Manufacturing Resources Planning (MRP II)</i>
1970s	<i>Material Requirements Planning (MRP)</i>
1960s	<i>Inventory Control Packages</i>

Tabela 2.1.4 - Evolução histórica; Fonte: Adaptado de Hossain, L.; & Rashid, M. A., (2002)

Do ano 2000 para cá, houve um impacto tecnológico muito progressivo, com a maior contribuição da internet, possibilitando o acesso aos sistemas a qualquer hora e em qualquer lugar. O que encurtou a distância entre fornecedor, produtor, revendedor e consumidor.

2.2. INFORMAÇÃO NAS ORGANIZAÇÕES

A existência de qualquer organização tem como missão a evolução, quando essa missão for bem acompanhada e monitorada com os sistemas de informação, o seu crescimento será rápido e progressivo. “As organizações dependem da utilização da informação, principalmente para apoio à tomada de decisões. O processo de tomada de decisão consiste na escolha de uma opção, entre diversas alternativas existentes, seguindo determinados passos previamente estabelecidos que culminam na resolução de um problema” (Carlos, F. M. M., 2008, p.17).

“De um modo geral, o ser humano, enquanto indivíduo, consome informação para tomar decisões e, desta forma concretizar em ação as suas intenções” (Gouveia, 2004).

“O dia-a-dia das organizações é também moldado em função da tecnologia disponível. Ao pensar nas organizações é necessário de alguma forma, ter em linha de conta esta realidade” (Teixeira, P. A. M.S., 2005, p.5).

“Computador insere um novo elemento na relação do indivíduo com a informação. Funciona como um dispositivo de mediação de comunicação à qual o indivíduo recorre para lidar com a informação quer para seu próprio uso, quer para interagir com outros indivíduos” (Carlos, F. M. M., 2008, p.18).

“A forma como esta informação é tratada na organização, é um dos fatores determinantes para o correto funcionamento” (Carlos, F. M. M., 2008, p.18).

O autor Laudon (2003) no livro “*Management Information Systems*” divide SI em quatro níveis, como mostra a figura 2.2.1, os quatro níveis estão atribuídos em seis tipos de acordo com a função de cada grupo dentro da organização.

Nível estratégico, *Executive Support Systems* (ESS) – Sistemas personalizados dirigidos à gestão de topo, bastante interativos e flexíveis que recebem

informação agregada e permitem realizar projeções para suporte à gestão da organização.

Nível intermédio ou de Gestão, *Management Information Systems (MIS)* – Sistemas direcionados para gestores intermédios que utilizam modelos simples para o tratamento de grandes volumes de informação, resultado das transações, e produzem relatórios síntese para a gestão da organização.

Decision Support Systems (DSS) – sistemas que ajudam os utilizadores na tomada de decisões não estruturáveis através de um processo interativo onde o sistema fornece informação, modelos e ferramentas para analisar a informação.

Nível do conhecimento, *Knowledge Work Systems (KWS)* – Sistemas utilizados por pessoal técnico, que incorporam modelos específicos de apoio ao diagnóstico e avaliação de situações complexas que requerem conhecimento especializado em áreas definidas.

Office Systems – Sistemas de suporte às atividades de comunicação e administrativas características de um ambiente de escritório.

Nível operacional, *Transaction Processing Systems (TPS)* – Sistemas informáticos que se caracterizam por desempenhar e registar as transações de rotina diárias necessárias à prossecução da atividade da organização, processam grandes quantidades de dados. A figura 2.2.1 ilustra estes tipos de sistemas da informação e sua hierarquia por nível organizacional.

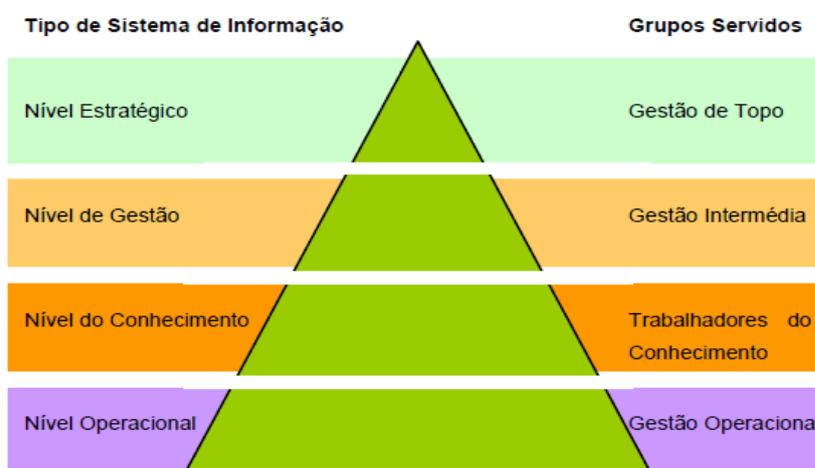


Figura 2.2.1 - Tipos de SI. Fonte: Adaptado de Laudon & Laudon, (2003)

“O sistema organizacional e os SI são indissociáveis. O conhecimento e compreensão de um deles em toda sua abrangência, implica intrinsecamente conhecer e entender completamente o outro” (Varajão, 2005). A figura 2.1.2 reforça as características da informação de tomada de decisão nas organizações. “A informação, bem como a sua recolha, tratamento e apresentação tem como objetivo final facilitar a tomada de decisões no negócio quer seja de nível operacional, quer tático, quer mesmo estratégico” (Carlos, F. M. M., 2008, p. 20).

A informação pode ser estruturada quando é usada no nível operacional, semiestruturada quando usada no nível tático e não estruturada quando usada na tomada de decisão ao nível estratégico ou topo da organização. “As organizações que recorrem o uso de sistemas de informação e tecnologias de comunicação devem considerar cinco elementos básicos: objetivos de negócio, *hardware*, *software*, procedimentos e pessoas” (Gouveia, 2004). A garantia de um bom enquadramento dos cinco elementos na organização, é assegurada pelo fluxo de dados, pela informação e pelo conhecimento dos indivíduos envolvidos na tomada de decisão na organização.



Figura 2.2.2 - Características da decisão; Fonte: Adaptado de Kimball (2002)

“Um sistema empresarial melhora o funcionamento interno da organização, o relacionamento com os clientes, o bom desempenho, e alcance de grandes descidas nos custos, e tudo isto, depende de como é implementado e adequado na organização” (Kimball, 2002). Pode-se observar que os méritos dos sistemas não são adquiridos por si só, mas também como são implementados na organização.

Uma informação é considerada útil e valiosa, quando responde atempadamente e de maneira segura aos objetivos para a qual foi concebida.

2.3. SISTEMAS OPERACIONAIS (OLTP) E SISTEMAS DE APOIO À DECISÃO (SAD).

Os sistemas operacionais são aqueles que são concebidos para otimizar a entrada de dados com estruturas complexas, enquanto os *Decision Support Systems* (DSS), ou sistemas de apoio à decisão (SAD) são utilizados na resolução de problemas complexos e não estruturados.

O funcionamento dos SAD combina técnicas analíticas com funções consideradas tradicionais no processamento de dados, como forma de recuperação da informação com interface interativo. “O modelo dos sistemas de suporte a decisão, e o de entidade-relacionamento, procura eliminar a redundância dos dados e facilitar a entrada dos mesmos. Mas cria muitas tabelas, que por vezes são difíceis de compreender por apresentarem diagramas muito grandes, o que também dificulta a navegação dos utilizadores de dados” (Kimball, 2002).

“Dado que normalmente existe uma BD operacional que suporta a atividade operacional da organização, podia colocar-se a hipótese de efetuar as pesquisas analíticas sobre esta mesma BD” Carlos, F. M. M. (2008, p.25). Segundo este mesmo autor, “as pesquisas são feitas através de sistemas *Online Transaction Processing* (OLTP), ou seja Sistemas de Processamento de Transações Online”, que poderá não ter o resultado mais desejado e originar uma das seguintes situações:

1 - Incapacidade de integração - O sistema estar definido a um nível de detalhe diferente daquele que é requerido para a plataforma analítica.

2 - Comprometimento do desempenho do sistema - Otimizado para dar resposta a um conjunto de interrogações e que passa a ter de lidar com questões analíticas *ad hoc*.

3 - Incapacidade – Não conseguir lidar com informações provenientes de outras fontes de dados.

Devido aos fatores acima referidos, antes da proliferação de computadores nas organizações, a análise de dados nos sistemas operacionais passava pelo sector informático que criava relatórios de auxílio. Com a proliferação de computadores a

análise passou a ser feita diretamente pelo utilizador que toma a decisão, correndo maior risco de duplicação e fraca qualidade de dados. “Os dados podem não ser extraídos no mesmo instante, e os algoritmos usados podem ser diferentes e os nomes de atributos podem também ter significados diferentes” (Kimball, 2002).

“Mesmo que uma organização detenha uma estrutura que contenha informação relevante no contexto da organização, existem várias razões que demonstram a dificuldade em efetuar uma solução analítica sobre esta estrutura” (Carlos, F. M. M., 2008, p.26). Ao tomar em consideração as afirmações acima citadas, a tabela 2.3.1 mostra o comportamento dos dados para a tomada de decisão usando OLTP, na visão de Cortes e Kimball.

CORTES	KIMBALL
Por norma a informação operacional está muito mais detalhada do que a utilizada para as funções analíticas.	Os sistemas operacionais geralmente estão dispersos por toda a organização e foram desenvolvidos ou adquiridos de forma independente ao longo do tempo, não possuindo a integração desejada.
A informação é atualizada várias vezes num sistema operacional o que dificulta que os dados funcionem de forma imutável num dado período de tempo como seria desejado para um sistema analítico.	Os dados sofrem alterações constantemente, tornando difícil a repetição de uma operação que forneça as mesmas informações.
Enquanto nos sistemas operacionais os utilizadores são basicamente técnicos com funções limitadas no âmbito das suas competências profissionais, no sistema analítico estão os gestores com permissão para cruzar vários sectores operacionais.	Tentar extrair informações com base no mesmo sistema, degrada a performance das aplicações rotineiras.
Enquanto no sistema operacional	A discrepância entre as informações

predominam as instruções de modificação de dados, nos sistemas analíticos as consultas, não sendo eficiente um sistema que contemple os dois tipos de operações.	provenientes de dois sistemas que usam aplicações diferentes pode levar a uma tomada de decisão baseada em factos não reais.
Normalmente os sistemas analíticos operam sobre uma maior quantidade de dados face aos sistemas operacionais que originam uma diferente necessidade de requisitos.	Os mesmos dados estão em inúmeras BD diferentes, isto significa que os mesmos dados são mantidos por diversas pessoas, sistemas e processos, consumindo recursos.

Tabela 2.3.1- Limitação de dados OLTP; Fonte: Adaptado Cortes e Kimball, em Carlos (2008)

“Os dados transacionais são usados pelos usuários em geral no dia-a-dia em seus processos e transações, gravação e leitura. Ex: consulta de estoque, registo de vendas” (Vaisman, 1998, p. 5). Uma das grandes vantagens desta ferramenta é a de reduzir ao mínimo o grau de redundância, mas é complexo para o utilizador final, na formulação de consultas devido à fragmentação dos dados em diversas tabelas. Após a verificação destes pontos, é de compreender a importância e necessidade da informação na organização assim como, a forma de armazenamento e manuseio da mesma.

2.4. CARACTERIZAÇÃO DE DATA WAREHOUSE

Caracterizar DW é descrever acima de tudo as propriedades que o constituem. Define-se um DW como sendo um repositório de dados, destinado a apoiar os gestores na tomada de decisão segura e inteligente. William H. Inmon (2002,p.31), considerado um dos pioneiros deste tipo tridimensional de armazenamento de dados, definiu DW como sendo “uma base de dados orientado por assunto, integrado, não volátil e histórico, concebida para suportar o processo de tomada de decisão”¹.

“O conteúdo do DW deve ser compreensível. Os dados devem ser intuitivos e evidentes para o utilizador empresarial, não apenas o desenvolvedor”² (Kimball, R. & Ross, M., 2002, p.2). Com a base das definições dos considerados pioneiros em

1 “A Data Warehouse is subject-oriented, integrated, nonvolatile, and time-variant collection of data in support of management’s decisions.”

2 “The contents of the data warehouse must be understandable. The data must be intuitive and obvious to the business user, not merely the developer.”

sistemas de armazenamento desta natureza, e com as realidades dos que já usam ou implementaram este tipo de repositório nas suas organizações, pode-se considerar DW como um instrumento chave na tomada de decisão segura e ágil.

“DW é uma ferramenta que oferece aos decisores de uma instituição ou empresa, o acesso a qualquer nível de informação, que possa apoiar na sobrevivência num mercado competitivo” (Singh, 2001, p.14). Ainda este autor, considera “Data Mart como subconjunto do DW que desempenha o papel de um DW departamental, regional ou funcional, podendo-se construir uma série deles ao longo do tempo e eventualmente vinculá-los através de um DW lógico para uma empresa completa”³.

“DW são aplicações construídas de raiz com o único propósito de servirem como base, ou como instrumento de apoio, à análise dos dados sobre os quais uma organização desenvolve as suas atividades” (Caldeira, C. P., 2008, p.19).

Ao fazer síntese das diversas definições e considerações dos autores de renome neste tipo de repositório de dados, verifica-se que a finalidade principal é o suporte estratégico no apoio à tomada de decisão.

DW é um sistema de armazenamento de dados muito complexo, que na sua conceção integra muitos componentes do tipo *hardware*, redes de computadores, software, sistemas e tecnologias de comunicação, servidores, bases de dados e sistemas operacionais.

Para o bom funcionamento deste repositório, devem reunir-se todos os dados ou toda a informação dispersa em sistemas operacionais da empresa, assim como, dados externos à empresa, mesmo em formatos diferentes. A figura 2.4.1 ilustra um modelo de Inmon de processo de dados de DW.

³ “Plays the role of a DW (departmental, regional or functional), and we could build a series of them over time and eventually link them through a company-whole logic DW”

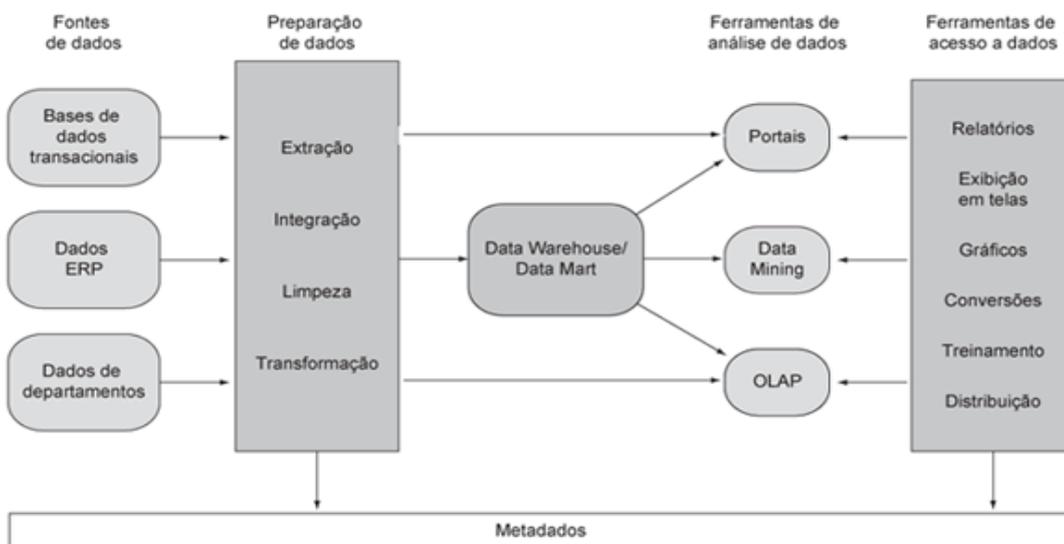


Figura 2.4.1 - Processo de data warehouse. Fonte: Adaptado de Inmon (1997).

Para melhor sustentabilidade dos dados, o DW necessita de uma boa arquitetura, hoje em dia, existem várias dessas arquiteturas, a figura 2.4.2 ilustra uma delas na visão de Inmon.

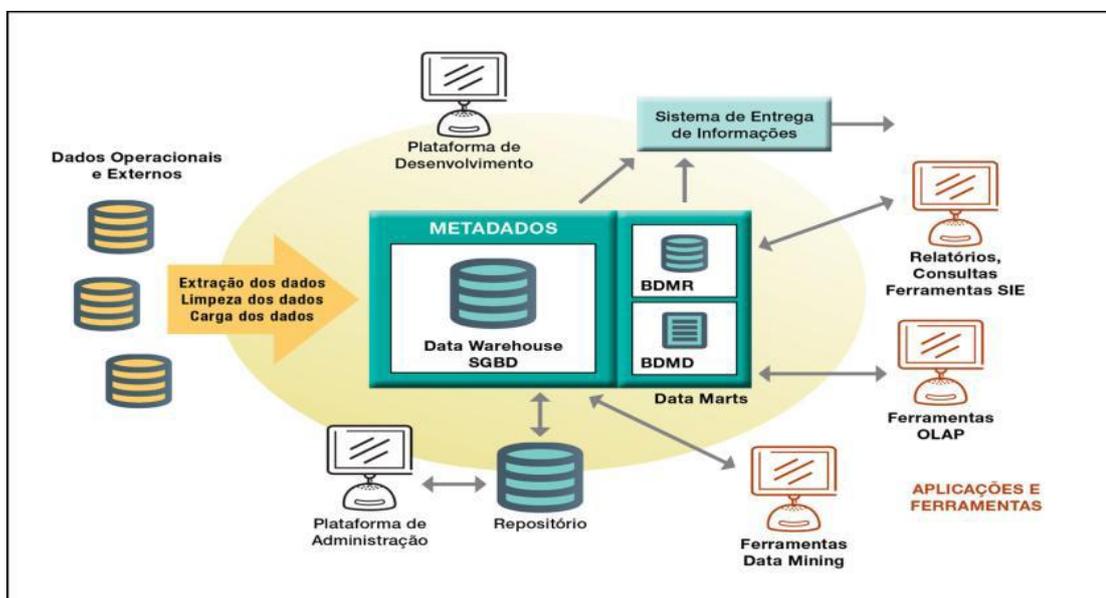


Figura 2.4.2 - Arquitetura de Data Warehouse; Fonte: Adaptado de Inmon (1997)

O processo de geração de informação para auxiliar na tomada de decisões não é uma atividade nova. No início de década 60, consistia na criação de aplicações individuais executadas sobre arquivos ou ficheiros mestres, o que criava redundâncias

nos dados, tornava difícil o controlo e originava problemas de baixa produtividade, falta de credibilidade dos dados e a impossibilidade de transformar dados em informações.

Com a aplicação de Sistemas de Gestão de Bases de Dados (SGBD), os dados passaram a ser integrados numa única fonte para todo o processamento, permitindo assim o aparecimento dos primeiros Sistemas de Apoio à Decisão (SAD). “O novo e crescente ambiente de negócio obrigavam cada vez mais à necessidade de separação do ambiente de processamento operacional do analítico” (Inmon, 1997). Com esta separação de ambientes, em 1992 surgiu o conceito de Data Warehouse.

Numa organização, o DW afasta-se dos planos estritamente operacionais e logísticos, porque este situa-se no plano mais elevado da ação estratégica e apresenta características específicas no mundo dos sistemas de informação.

Facundes, E. M. (SD) considera:

A principal proposta do *data warehouse* é pôr nas mãos dos gestores de negócios dados estratégicos propícios para as tomadas de decisões baseadas em fatos verídicos e não por intuição. A produtividade oferecida pelo *data warehouse* é traduzida em ganho de tempo e dinheiro. Com o cruzamento de dados, seleciona-se informação que os sistemas tradicionais ou leigos não conseguem identificar.

Em determinadas empresas quando um gestor estratégico faz uma pergunta sobre o perfil dos seus negócios fora de um padrão definido pelos analistas de sistemas o tempo e o esforço para responder é longo. Porque várias bases de dados devem ser consultadas para manipulação dos dados e gerar relatórios e gráficos.

Se uma decisão deve ser tomada acompanhando a velocidade do mercado competitivo, o executivo acaba tomando a decisão por meios alternativos não podendo esperar o resultado da área de sistemas.

Quando o assunto é simulação de cenários de negócios o processo é ainda mais traumático. Os sistemas estruturados atuais não são suficientemente flexíveis para traçar cenários complexos usando cruzamento de vários componentes.

Como se pode verificar, o intuito não é o DW tomar decisões, mas sim proporcionar uma boa informação que informe melhor e apoie os decisores nas suas funções. Em DW os utilizadores são considerados apenas como consumidores da informação e conhecimentos. Face à interação homem – sistema – máquina, das ferramentas que potenciam o processamento analítico dos dados, o mais comum é o *Online Analytic Processing (OLAP)*.

Os sistemas de DW assumem o papel de concentradores dos vários dados dispersos pela organização, armazenando de uma forma integrada e homogênea. “As principais características de um Data Warehouse/Data Mart são: facilidades de consulta da informação e consistência, adaptabilidade e flexibilidade às mudanças, proteção destas informações e utilização da informação como base de tomada de decisão” (Kimball, R.; Ross, M., 2002, p. 494). A figura 2.4.3 mostra os quatro componentes distintos e separados “Sistemas operacionais, *data staging area*, Área de apresentação de dados e Ferramentas de acesso a dados” existentes no ambiente DW.

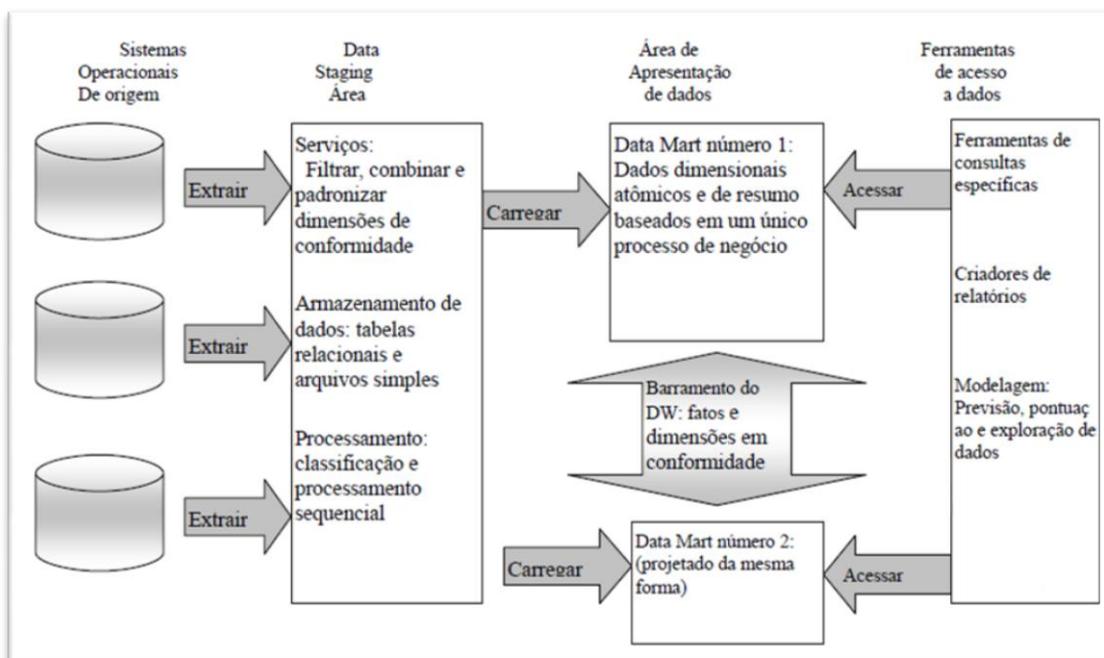


Figura 2.4.3 - Elementos básicos de DW; Fonte: Kimball, R., Ross, M. (2002, p. 494)

Cada um dos quatro elementos da figura 2.4.3 tem um papel determinante no bom funcionamento de DW, como se pode ver a seguir:

1 - Sistemas operacionais de origem – desempenham o papel de captura das transações da instituição.

2 - *Data staging area* – local destinado a preparar dados e processos de origem que serão utilizados.

3 - Área de apresentação de dados – onde ficam armazenados os dados disponibilizados ao utilizador final.

4 - Ferramentas de acesso aos dados – ferramentas OLAP e de mineração de dados que facultam ao utilizador o manuseio e uso de dados de forma mais rápida, interativa, para executar análises mensuráveis. São estas que asseguram a consistência, integração e precisão, usando os *Data Marts* como fonte de dados. “Independentemente do suporte físico dos dados, (em estruturas relacionais para microdados e multidimensionais para macrodados agregados) várias são as potenciais formas de visualização da informação contida no DW” (Quaresma, S., 2004).

2.5. MODELOS E TABELAS DE DATA WAREHOUSE

Ao falar dos modelos de um DW, é necessário ter-se em consideração antes de tudo, os elementos básicos que compõem esses modelos. Por isso, de forma resumida vai-se descrever os dois formatos de tabelas de DW.

Tabelas de facto – São normalmente usadas em modelo estrela, são normalizadas, não contém redundâncias, e é neste tipo de tabela que são representados os eventos de negócio da empresa, permitindo a análise dos processos desses eventos.

Tabelas de Dimensão – São as que rodeiam ou alimentam a tabela de facto, geralmente são não normalizadas, contendo por vezes muita redundância nos dados. São responsáveis pela visualização de dados, classificação e descrição dos elementos participativos nas tabelas de facto é por isso que se considera não possuir atributos numéricos.

Não basta falar de tabelas e não focar o que as caracteriza. As tabelas são constituídas por variáveis. “Variáveis são atributos numéricos que representam um acontecimento ou evento, pela combinação das participações de tabelas dimensionais” (Machado, 2006), como pode-se ver na figura 2.5, a tabela de facto “Vendas acumuladas” com atributos e em sua volta as tabelas de dimensão.



Figura 2.5 - Dimensões e de facto com os seus atributos. Fonte: Scalabrin, E. E. (ND)

2.6. MODELOS DE DW

Existem vários modelos em DW, mas na essência os principais são três “estrela, floco de neve, e constelação”. Na generalidade basta que se opte por um destes três esquemas de BD, para que se configure um DW.

2.6.1. Modelo em Estrela

Este modelo é vulgarmente conhecido como *Star Schema*. É assim denominado devido a sua constituição em forma de estrela, estando ao centro a tabela de facto e rodeada por tabelas de dimensão, como ilustra a figura 2.5 acima. As tabelas dimensionais podem ser normalizadas ou não, mas o mais prático é não serem, visto que recebem dados vindos de muitos pontos diferentes. Cada tabela de dimensão tem apenas uma ligação com a tabela de facto fortemente normalizada.

A tabela de facto pode estar ligada a várias tabelas de dimensão. Este modelo em estrela foi criado pelo americano Ralph Kimball em 1998 e é caracterizado pela presença considerável de dados redundantes nas tabelas dimensionais.

O melhor funcionamento e rapidez deste modelo é assegurado na medida em que as pesquisas são feitas primeiramente nas tabelas dimensionais, por intermédio de uma estrutura de chaves, para que não seja necessário percorrerem todas as tabelas.

A principal propriedade deste modelo é de possuir uma única tabela de facto contendo dados não redundantes. Cada chave primária na tabela de facto corresponde a uma de dimensão.

Tem como vantagem, o fato de disponibilizar ao utilizador final toda a informação necessária para a sua tomada de decisão segura e rápida, facilitar a incorporação de novos módulos e respetiva manutenção.

A desvantagem é da tabela de facto depender das tabelas de dimensão. Como se viu anteriormente as tabelas de dimensão não são normalizadas, o que permite repetições de informação. As tabelas de dimensão por não serem normalizadas, ao alterar um determinado dado numa das tabelas, o sistema não atualiza automaticamente em todas as outras tabelas onde este dado é mencionado. A figura 2.6.1 representa um modelo em estrela.

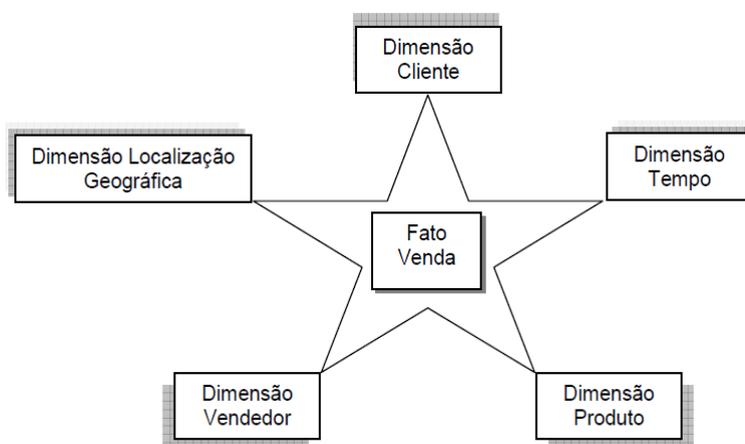


Figura 2.6.1 - Modelo estrela; Fonte: Adaptado de Machado (2006)

2.6.2. Modelo em Floco de Neve

Este modelo tem uma estrutura semelhante ao modelo em estrela, e é muito comum a sua aplicação. Possui a tabela de facto e tabelas dimensionais, com a particularidade do terminal ou extremidade de cada tabela dimensional, poder constituir o centro de outras estrelas. É este ponto que diferencia o modelo em estrela. A figura 2.6.2 com a tabela de facto “Vendas”, esquematiza a estrutura deste modelo.

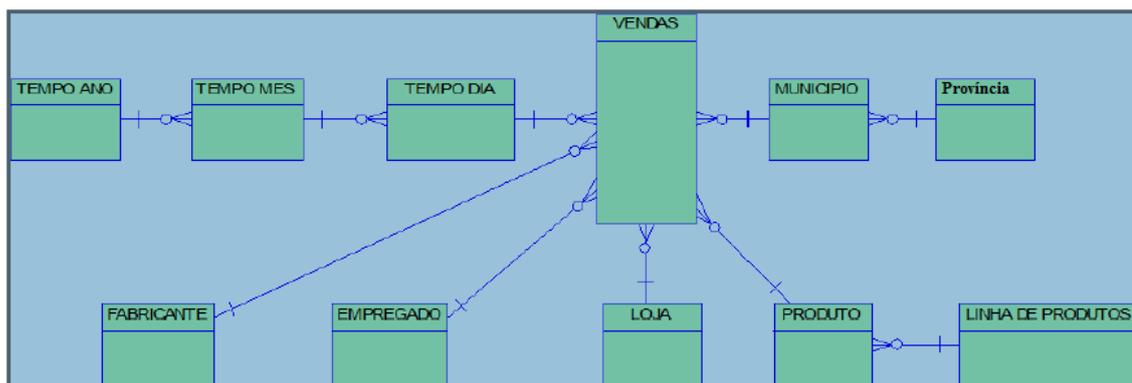


Figura 2.6.2 - Modelo Floco de Neve

Não é aconselhável a transição do modelo em estrela para este tipo, devido à grande complexidade no uso final para o utilizador.

Oferece vantagens em consultas nas estrelas com mais de uma dimensão em comum e conformes. É um modelo normalizado, “evita assim logo de princípio as redundâncias nos valores em tabelas” (Machado, 2006).

2.6.3. Modelo em Constelação

É um modelo que pode ser visto como uma junção de muitos modelos em estrela, por terem muitas tabelas de facto e dimensionais que partilham entre si. Alguns projetistas consideram que os Data Marts utilizam o modelo em estrela enquanto os *Data warehouses* utilizam o modelo em constelação. A figura 2.6.3 representa este modelo.

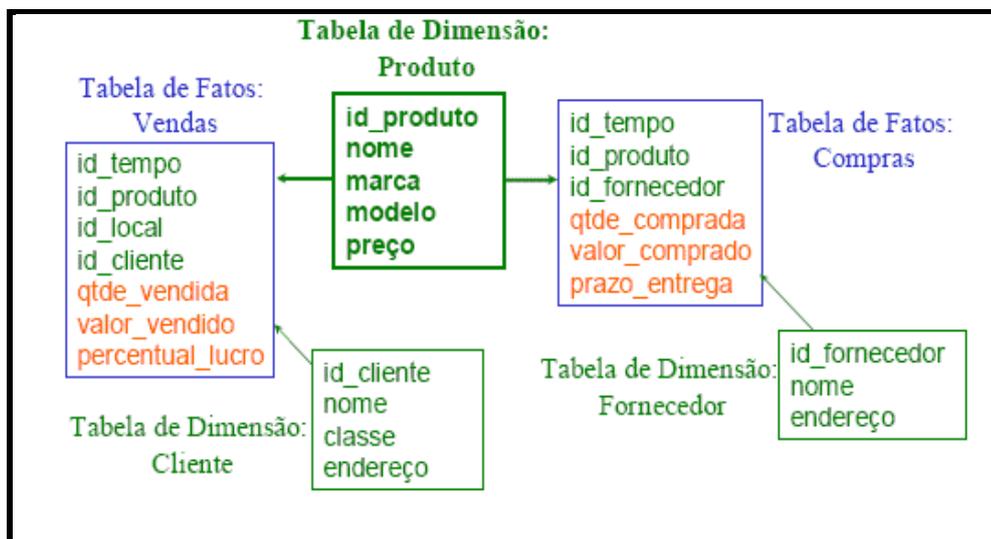


Figura 2.6.3 - Modelo em constelação; Fonte: Carlos, F. M. M. (2008)

2.7. PROCESSO DE MODELAÇÃO

É uma ação onde se define as tabelas de facto e dimensão, atributos, relacionamento e hierarquia. “Processo de modelação dimensional adapta-se na perspetiva do utilizador final, escolher as dimensões e os factos a incluir no DW” (Carlos, F. M. M., 2008, p.40). Considera-se ainda que o processo de modelação é *top down* e que se deve ter em conta os aspetos mencionados na tabela 2.7

Processo de negócio	É necessário responder à questão: onde e como são recolhidos os dados?
Granularidade	A granularidade determina a dimensionalidade do DW e tem impacto no seu tamanho. Quase sempre faz sentido guardar os dados acerca das dimensões com a maior granularidade possível. O objetivo não é ver cada registo individualmente, mas permitir que as pesquisas sejam mais precisas.
Dimensões da tabela de factos	Geralmente é possível acrescentar outras dimensões, desde que exista apenas um valor dessas dimensões para cada combinação de valores das dimensões já existentes. Se isso não acontecer, é necessário rever a granularidade do modelo.
Medidas na tabela de factos	Na maioria dos casos é uma perda de tempo tentar normalizar as tabelas de dimensão pois iria dificultar a atividade em que o utilizador explora uma única dimensão com o objetivo de definir colunas e restrições para uma pesquisa posterior que consiste em verificar o valor de determinados atributos quando se restringe o valor de outros atributos.
Dimensão tempo	A dimensão tempo, presente em quase todos os DW apresenta vantagens em relação à data em SQL como: permitir análise de dias da semana/fins-de-semana; facilita a divisão em períodos fiscais; permite análise de vendas em feriados e datas especiais; cada registo na tabela representa um dia; a tabela não é construída a partir dos sistemas fonte/operacionais; pode conter dias que ainda não aconteceram e conter campos que permitem análise de dados em períodos temporais.
Hierarquias explícitas	A definição de hierarquias explícitas facilita as operações de <i>drill-down</i> e <i>rollup</i> , ou seja, em aumentar e diminuir o nível de

	detalhe de uma consulta.
Tabela de factos sem factos	Existem no entanto, alguns processos de negócio suscetíveis de serem modelados num DW, aos quais não existem factos associados. Nestas situações utilizam-se tabelas de factos sem factos.
Dimensões muito grandes	<p>Este tipo de dimensões necessita de um tratamento diferente de forma a acelerar as pesquisas e a facilitar alterações. Algumas das soluções apontadas passam pela criação de índices apenas nos campos utilizados para fazer pesquisas e na criação de mini dimensões com campos que mudam com frequência. Os campos demográficos são bastante utilizados, quer individualmente, quer em conjunto, para restringir as consultas num DW. A forma mais eficaz de utilizar estes atributos consiste em colocá-los numa mini-dimensão separada.</p> <p>Para atributos com carácter contínuo devem utilizar-se gama de valores e para dimensões muito grandes, as mini-dimensões permitem poupar espaço e acelerar as consultas.</p>

Tabela 2.7 - Aspectos no Processo de Modelação; Fonte: Adaptado de Kimball (2002)

2.8. ERROS MAIS FREQUENTES NA MODELAÇÃO.

Em qualquer modelação de dados existem erros, sendo este trabalho um modelo conceptual, logicamente exigirá modelação de dados, sendo assim, é sempre bom precaver os possíveis erros que surgem com maior frequência e que facilmente são cometidos com maior incidência. Kimball (2002), citado em Carlos, F. M. M. (2008, p.42) considera de relevo os seguintes erros:

1- Tabelas de facto com atributos de texto

Os atributos de texto não podem estar na tabela de factos, esta tabela deve conter apenas as medidas numéricas. As descrições dos atributos devem ser colocadas nas tabelas de dimensão.

2- Atributos com descrição limitada

Geralmente a tabela de factos é maior em relação as dimensionais, o que leva muitas vezes à limitação na descrição dos contextos nestas tabelas. É de salientar que é necessária a descrição completa por cada código da tabela dimensão.

3- Hierarquias em múltiplas dimensões

As hierarquias devem estar em forma de cascata de relacionamento de muitos para 1, pertencendo a uma única tabela de dimensão.

4- Alterações nas dimensões

As alterações dos atributos nas dimensões não podem ser ignoradas, devido à dinâmica constante do mercado.

5- Problemas com hardware

Faz parte do programa de manutenção e evolução dos sistemas de suporte à decisão, ao adicionar mais *hardware* para o melhoramento do desempenho.

6- Chaves operacionais e primárias

As chaves operacionais não podem e nem devem ser usadas como chaves primárias em tabelas de dimensão, porque pode trazer sérios problemas no Data Warehouse.

7- Granularidade da tabela de factos

Não se deve negligenciar a granularidade da tabela de factos, o passo primário é a seleção dos processos para a tabela de factos. Em seguida, definir com exatidão a granularidade dos dados, para construção da tabela de factos e escolha das dimensões validas para a granularidade desejada.

8- Modelo dimensional com base num relatório

A tabela de factos deve representar os processos de negócio da empresa e não apenas para satisfazer um determinado requisito, por isso, as tabelas dimensionais não devem constituir um determinado relatório de negócios.

9- Pesquisa de dados atômicos num formato normalizado

Os modelos normalizados são úteis de certa forma, mas não devem ser usados para fornecer dados aos utilizadores, com isto, significa que os utilizadores nunca devem pesquisar dados atômicos num formato de modelo normalizado, visto que, dados atômicos são a fundação do modelo dimensional.

10- Conformidade nas tabelas de factos e dimensões

Em processos diferentes de negócios, quanto menos *Data Marts* isolados melhor. Se as tabelas de facto associadas são da mesma dimensão então as tabelas de dimensão devem ser idênticas. O segredo na construção de *Data Marts* distribuídos

está nas tabelas dimensionais, o que permite a inclusão de novas fontes e funcionamento de múltiplas tecnologias não compatíveis.

2.9. FERRAMENTAS DE BUSCA, TRANSFORMAÇÃO E ACESSO

Para que se possa ter um bom repositório de dados com qualidade desejável num mercado competitivo, é necessário que o DW esteja provido de ferramentas tecnicamente aceitáveis em ambientes multidimensionais. Não se irá descrever na íntegra todas as ferramentas disponíveis no mercado, apenas referenciar e descrever de forma resumida a ETL e OLAP, uma vez que, OLTP já foi referenciada acima nos sistemas operacionais. A tabela 2.9.1 referencia as ferramentas e sua finalidade.

Ferramenta	Finalidade e breve descrição
Armazenamento de dados.	Considerada como tendo grande potencial, por ser este um dos grandes objetivos da existência do DW.
Extração de dados	Responsável pela busca nas bases de dados operacionais, os dados que serão armazenados em DW. Esta ferramenta está diretamente relacionada com os metadados (MD) por estes permitirem saber a origem da informação, as alterações, tendo como função principal, manter a consistência dos dados.
Transferência de dados e replicação	Considerado subconjunto das ferramentas de extração. Não transforma nem processa dados, apenas muda os dados de um lugar para outro, usada geralmente na consulta ou pesquisa, para facilitar ou dar resposta às buscas de uma forma mais rápida, agilizando o serviço solicitado.
Transformação	É através desta que os dados em formatos diferentes, são ajustados e transformados para o formato do DW. Estando já em formato de DW, as pesquisas ou consultas ficam facilitadas.
Limpeza de dados	Como próprio nome diz, limpa os dados, ajusta-os, faz as devidas correções e fusões de dados quando necessário.
Gestão	Monitoriza a performance e segurança do sistema

Consultas	Usadas para consultas ou gerar relatórios com os dados retirados de DW, apresentando em formato apropriado
Gestão de relatórios	Esta ferramenta é semelhante à de consultas, só que ela é direcionada para os relatórios mais complexos, por exemplo aqueles que contém gráficos, sejam eles relatórios sintéticos ou analíticos.
Data mining	É utilizada na descoberta de novas tendências entre as informações da empresa, novos padrões e novas correlações, através da análise de grandes quantidades de dados armazenados em DW usando técnicas estatísticas e matemáticas apropriadas.
Simulação	Simula cenários respondendo a perguntas do tipo “e se”.
OLAP	É a parte mais visível do DW, e faz a análise dos dados. Auxiliam os gestores a sintetizarem as informações das organizações por meio de comparações, análise histórica e projeções.

Tabela 2.9.1 - Ferramentas de DW; Fonte: Adaptado de Dill, S. L. (2002)

2.9.1. ETL

É uma ferramenta com a função de extrair dados heterogêneos na base de dados operacionais ou outras fontes, transformar, limpar e carregar no DW ou DM segundo as necessidades da empresa no mercado. Muitos autores consideram esta como sendo a fase mais crítica e demorada na construção de DW ou do posterior DM.

A tomada de decisões com apoio nos dados armazenados em DW provém da informação fornecida pelas ferramentas de tipo *front-end*, “se estes dados não forem corretamente trabalhados no processo de extração, as informações geradas através deles farão com que decisões sejam tomadas erroneamente, podendo afetar diretamente os negócios da organização” (Abreu, F. S. G.G., 2008).

“A extração de dados pode levar aproximadamente 60% de horas de construção de DW” (Kimball, 1998). Este processo deve basear-se na busca em bases tradicionais ou externos, informação importante que esteja de acordo com a

modelagem do sistema. “No momento de criação do DW é normal uma carga inicial de dados que faça com que a extração vá buscar todos os dados dos sistemas fontes em uso na organização ou específicos de determinados assuntos. Mas com o andar do tempo a extração deve estar preparada para fazer somente carregamentos incrementais. A carga incremental que carregue somente os registros que foram alterados ou inseridos desde a carga inicial é muito mais eficiente” (Kimball, 1998, citado em Abreu, F. S. G. G., 2008).

“A transformação dos dados é a fase subsequente à sua extração. Esta fase não só transforma os dados, mas também realiza a limpeza dos mesmos. A correção de erros de digitação, a descoberta de violações de integridade, a substituição de caracteres” (Gonçalves, 2003) citado em (Abreu, F. S. G. G., 2008).

As afirmações destes autores confirmam claramente que a fase de ETL é a mais crítica na implementação de DW. “Um exemplo bem comum de utilização do ETL para sistemas transacionais é o uso de informações de bases corporativas para estes sistemas” (Abreu, F. S. G. G., 2008).

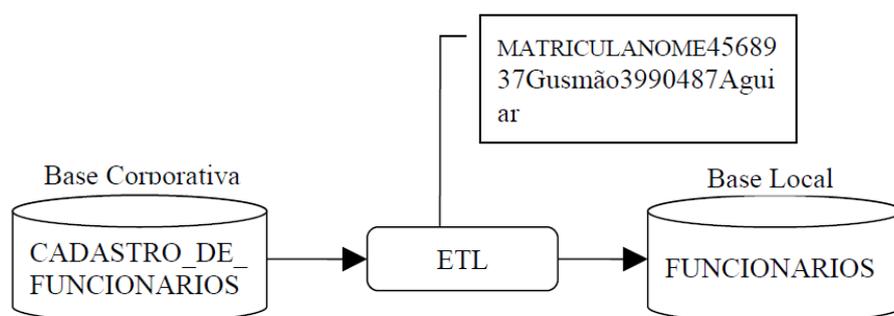


Figura 2.9.1 - Processo ETL; fonte: Abreu, F. S. G. G. (2008)

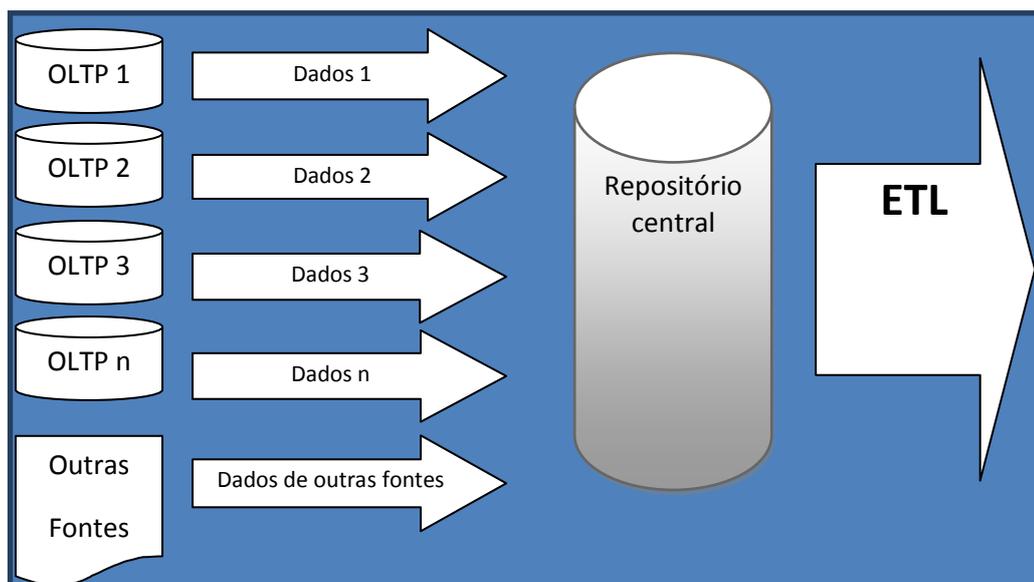


Figura 2.9.2 – Carregamento e transformação de dados

2.9.2. Ferramenta OLAP

Esta é a ferramenta mais popular em termos de uso na exploração em tempo real de um DW. É uma ferramenta de *Business Intelligence*, muito utilizada principalmente em modelos de DW em estrela, e também se presta a análises *ad hoc*. Possui uma grande capacidade para efetuar cálculos considerados complexos e para a conversão de volumosas bases de dados em informação.

“O termo OLAP é sinónimo de conjunto de ferramentas direcionadas ao acesso e análise *ad hoc*, com a finalidade de transformar dados em informações capazes de apoiar aos gestores na tomada de decisões de forma amigável e flexível em tempo hábil” (Araújo, E. M. T., SD). “As ferramentas OLAP proporcionam condições de análise de dados *on-line* necessários para responder as possíveis perguntas dos analistas, gerentes e executivos” (Machado, 2004) citado em Araújo, E. M. T. (ND).

Ferramentas OLAP são aplicações de apoio ao utilizador final, para conceber ou consultar relatórios com a informação extraída em DW, para apoio à tomada de decisão. Em DWBRASIL OLAP (2005) citado em Araújo, E. M. T. (SD) no seu artigo “um estudo sobre as ferramentas OLAP”, afirma que as características desta ferramenta “é a capacidade de efetuar algumas operações” tais como:

Drill Across: ocorre quando o usuário salta um nível intermediário dentro de uma mesma dimensão. Por exemplo, a dimensão tempo é composta por ano, semestre, trimestre, mês e dia. A operação *Drill Across* é executada quando o usuário passa de ano direto para trimestre ou mês.

Drill Down: acontece quando o utilizador aumenta o nível de detalhe da informação, diminuindo a granularidade.

Drill Up: é o contrário do *Drill Down*, ocorre quando o usuário aumenta a granularidade, diminuindo o nível de detalhe da informação;

Drill Through: ocorre quando o usuário passa de uma informação contida numa dimensão para uma outra. Por exemplo: Inicia na dimensão do tempo e no próximo passo analisa a informação por região;

Slice and Dice: é uma das principais características de uma ferramenta OLAP. Como a ferramenta OLAP recupera o microcubo (No OLAP, as informações são armazenadas em cubos multidimensionais, que gravam valores quantitativos e medidas, permitindo visualização através de diversos ângulos. Estas medidas são organizadas em categorias descritivas, chamadas de dimensões e formam, assim, a estrutura do cubo), surgiu a necessidade de criar um módulo, que se convencionou de *Slice and Dice*, para ficar responsável por trabalhar esta informação. Serve para modificar a posição de uma informação, trocar linhas por colunas de maneira a facilitar a compreensão dos utilizadores e girar o cubo sempre que houver necessidade.

As ferramentas OLAP obedecem a uma estrutura cliente /Servidor multiutilizadores e podem ser classificadas em cinco seguintes tipos (Inmon, 1997):

MOLAP (*Multidimensional Online Analytical processing*);

ROLAP (*Relational Online Processing*);

HOLAP (*Hybrid Online Analytical Processing*);

DOLAP (*Desktop Online Analytical Processing*);

WOLAP (*Web Online Analytical Processing*).

Para armazenamento de dados esta ferramenta obedece a vários ambientes, no caso de Rolap e Molap obedece ao ambiente de dados multidimensional, enquanto que Holap, Dolap e Wolap podem estar em ambientes locais de bases de dados,

ficheiros ou em memórias de *workstations*. Claramente se vê que a principal característica de OLAP é facilitar ao utilizador final as consultas e construção de relatórios, mesmo que o utilizador não tenha bases sólidas em análises apoiadas nos sistemas e tecnologias da informação. A tabela 2.9.2 faz uma breve comparação das ferramentas OLTP e OLAP.

CARACTERÍSTICAS	Base Dados Transaccional	Base Dados Multidimensional
Finalidade	Registo de operações diárias	Análise do negócio
Uso	Operacional	Informativo
Tipo de processamento	OLTP	OLAP
Unidade de trabalho	Inserir, alterar, eliminar, consultar	Carregar, consultar
Número de utilizadores	Milhares	Centenas
Tipo de utilizador	Operadores	Decisores
Condições dos dados	Dados operacionais	Dados Analíticos
Volume	Megabytes - gigabytes	Gigabytes - terabytes
Histórico	60 a 90 dias	5 a 10 anos
Granularidade	Detalhados	Detalhados e resumidos
Redundância	Não ocorre	Ocorre
Estrutura	Estática	Variável
Manutenção desejada	Mínima	Constante
Acesso a registos	Dezenas	Milhares
Actualização	Contínua (tempo real)	Periódica (em <i>batch</i>)
Integridade	Transacção	A cada actualização

Tabela 2.9.2 - Ferramentas OLTP e OLAP; Fonte: Silva, C. I. M. e (2011)

2.10. BUSINESS INTELLIGENCE (BI) E DATA MINING

BI aqui descrito, é o incorporado na ferramenta SQL server 2008 da Microsoft, porque é esta ferramenta a proposta para construção do DW para INE-M. “O *Business Intelligence* consiste num conjunto vasto de aplicações e tecnologias para recolha, armazenamento, análise e acesso a dados que ajudem os gestores a tomar melhores decisões. Inclui, entre outros, sistemas para apoio à decisão, análise estatística e o *Data Mining (dmg)*” (Crisp, 2007).

Tomando como base a estrutura organizacional de um SI, o BI comporta os níveis: Operacional, destinada à automação das decisões; Tático, direcionado ao apoio na tomada de decisão e nível Estratégico, ao apoio à decisão operacional e sistemas de apoio ao executivo de topo.

“*Data Mining* é o uso de muitos métodos de aprendizagem e análise para extração de conhecimentos através de dados, passando por diversas fases como: preparação dos dados, compreensão do processo do negócio, avaliação dos modelos, modelação dos algoritmos, modelação das técnicas e disponibilização na produção” (Crisp, 2007). Com vista a solucionar problemas e tarefas em *Data Mining*, são utilizados algoritmos específicos como:

- a) **Métodos estatísticos** – destinados a problemas de predição e classificação.
- b) **Árvores de decisão** – aplicadas em tarefas de classificação.
- c) **Regras de associação** – usadas em *Market Basket Analysis*.
- d) **Clustering** – aplica-se no agrupamento de dados.
- e) **Técnicas de Visualização** – usadas em 4D ou mais dimensões com coordenadas paralelas e gráficos de dispersão.

No âmbito do BI para este estudo, a ferramenta SQL server 2008 possui muita potencialidade para conceção do projeto de DW. Na visão de Silva, C. I. M e (2011, p.60) destaca “*Analysis Services, Business Intelligence Development Studio, Integration Services e Reporting Services*” como pontos fortes de SQL server 2008 na construção de DW. Ainda segundo Silva, C. I. M e (2011, p. 60) “para além destes recursos, o *Microsoft SQL Server 2008*, fornece ainda dois ambientes distintos para a construção, implementação e gestão do recurso de *Analysis Services: O Business Intelligence Development Studio* e o *SQL Server Management Studio*”. Na utilização desta ferramenta, o autor aconselha sempre que possível acoplar na mesma plataforma de BI duas ferramentas de construção de relatórios, o *Excel Services* e o *Report Builder 2.0* para apoiar o utilizador final no manuseio de relatórios existentes, ou ter permissão para elaborar novos relatórios.

O *Excel Services* pertence ao *Microsoft Office SharePoint Server 2007*, usa folhas de cálculo do Microsoft Excel 2007. A sua função é simplificar, partilhar, garantir a segurança e gestão dos relatórios interativos de forma consistente. A interação com o DW nesta plataforma é feita através da ferramenta colaborativa considerada *front-end*, denominada *Windows SharePoint Services 3.0*.

2.11. CUSTOS E BENEFÍCIOS DO DW

A construção ou montagem de qualquer estrutura envolve custos e benefícios, assim acontece com o DW. É aconselhável analisar os custos e benefícios muito antes de o construir, de forma que a organização ou empresa possa se organizar e prever os investimentos possíveis para que haja satisfação no uso de DW nos seus negócios.

2.11.1. Custos

É necessário organizar e prever quanto antes, os custos para implementar DW numa organização, para que possa atingir os objetivos para os quais é concebido. Caso contrário, poderão surgir problemas sérios que até possam impedir a sua construção. “A classificação dos custos decorrentes da implantação do DW consiste na divisão do investimento, entre custo inicial e custo periódico. Também pode ser dividido, quanto ao tipo de investimento, em custos capitais e custos operacionais” (Carlos, F. M. M., 2008, p.53).

“Os custos capitais são referentes à compra de equipamentos, serviços e estrutura. Enquanto os custos operacionais estão relacionados com pessoal, não necessariamente gastos financeiros, mas sim gastos de tempo para o desenvolvimento de soluções” (Inmon, 2002, p.66). Nesta ótica os custos podem ser divididos em inicial e periódicos, consoante a sua aplicabilidade e destino.

Custo Inicial corresponde ao investimento efetuado com o *hardware*, *software* e pessoal, aquando a implantação do DW. O custo periódico é aquele que é oriundo do uso de DW, atualização e manutenção. Pode igualmente corresponder ao *hardware*, *software* e pessoal. Os fatores que podem influir nos custos segundo Carlos, F. M. M. (2008, p.55) podem ser:

- 1- **A quantidade de dados gerados pelo dia-a-dia operacional** – determina o crescimento do equipamento e do *software* do DW.
- 2- **O histórico que deve ser armazenado no DW** – a quantidade de dados que poderá ser armazenada.
- 3- **O necessário nível de detalhe dos dados** – quanto maior o detalhe contido nos dados, mais espaço de armazenamento é necessário.

- 4- **O nível de formação** que os utilizadores necessitam para se adaptarem ao DW.
- 5- **Se está num mercado competitivo ou não** – que tipo de análise deve ser feita pelos programas que trabalham sobre o DW.
- 6- **A velocidade com que o DW precisa ser construído** – quanto mais rápido, mais profissionais trabalham e consomem tempo a desenvolvê-lo, e como consequência, maior deve ser o investimento.
- 7- **Se o DW será construído manualmente ou através de ferramentas automáticas** – o processo manual requer menos gasto com *software*, mas muito mais gastos com pessoal, tanto no custo inicial quanto no custo periódico.
- 8- **O número de *Data Marts* a construir** – quantos departamentos necessitam da sua própria visualização dos dados analíticos, tendo a necessidade de construir os seus próprios *Data Marts*.
- 9- **A quantidade de sumários e dados sumariados a ser mantido** – quanto mais sumários, menor a quantidade de acesso aos dados detalhados, que podem estar armazenados em dispositivos de armazenamento de massa.
- 10- **A escolha do SGBD correto**, que contenha o maior número de ferramentas de ETL e administração nativas.
- 11- **A escolha do *hardware* correto**, de acordo com as necessidades de processamento e armazenamento.
- 12- **A contratação de bons profissionais** para realizar cada tipo de serviço
- 13- **Erros na implementação** devido a “economias” na hora de fazer o projeto do DW.

2.11.2. Benefícios

Os benefícios de um DW na organização podem ser medidos de diversas formas, como por exemplo, um grande retorno com menor investimento, dá vantagens nos negócios, aumenta os lucros e a competitividade da organização. “O investimento num DW não compensa se o custo não for pago gradualmente com o uso

dele. Assim, para que o DW se pague mais facilmente, é preciso encontrar o equilíbrio entre o custo do DW e o benefício para a organização” (Carlos, F. M. M., 2008, p. 55).

“Dos vários benefícios que o DW traz ao cotidiano de uma organização, o mais simples de se ver é a facilidade de extração de relatórios integrados e sua diminuição de custo ao extrair relatórios sem um DW” (Carlos, F. M. M., 2008, p.55).

Na visão do autor Inmon (2008), uma organização que possui várias aplicações operacionais precisa de:

1. Localizar em que sistemas estão os dados;
2. Recuperar os dados;
3. Extrair dados de cada um dos sistemas operacionais;
4. Integrar/transformar os dados;
5. Padronizar os dados;
6. Reconciliar os dados;
7. Modificar as estruturas de dados;
8. Reunir os dados;
9. Construir o relatório.

Numa visão geral pode-se observar que o esforço na produção de relatórios operacionais envolve muitos custos e um tempo muito longo, com a particularidade de não se reaproveitar, visto que, as necessidades mudam consoante o tipo de relatório.

Usando DW, os relatórios que demoravam semanas e meses para estarem prontos, podem levar apenas horas ou alguns dias, porque alguns dos passos do modelo operacional, no DW são feitos num único passo.

2.12. FATORES CRÍTICOS DE SUCESSO DO DW

Consideram-se fatores críticos de sucesso do DW, os que determinam o êxito da sua implementação na organização. Perkins, A. (2000) indica alguns fatores críticos de sucesso:

1. Definição do projeto
2. Administração de expectativas
3. Patrocinador forte
4. Envolvimento dos utilizadores

5. Pessoal qualificado no projeto
6. Estratégia de implementação da arquitetura
7. Extração, transformação e carga dos dados
8. Modelo dos dados
9. Metadados
10. Conceção da Arquitetura e Infraestrutura Tecnológica
11. Ferramentas dos utilizadores finais
12. Acompanhamento

3. METODOLOGIA

Uma pesquisa é uma atividade sistemática realizada com base em certos métodos e técnicas, com o intuito de atingir um objetivo científico. Para o alcance desses objetivos não basta efetuar simplesmente a quantificação de fenômenos. Precisa de descrever e demonstrar tudo o que estiver em redor do cenário.

“O objetivo de estudar detalhadamente a complexidade do mundo cotidiano pode ser alcançado através de uma variedade de métodos. A atividade básica da ciência é a pesquisa, inserida como atividade crítica e criativa de questionamento sistemático” (Lustosa, R. B., 2009, p.92). Ainda este mesmo autor refere que “a ciência é um processo racional e permanente de busca da verdade ou de respostas a questões propostas. Para atingir uma verdade, é preciso seguir formalidades orientadas por princípios científicos estabelecidos com o objetivo de criar formas para validar o estudo científico”. Com isto realça-se que só com pesquisa científica podem atingir-se os resultados preconizados nos objetivos da investigação.

3.1. METODOLOGIA DE PESQUISA

Sendo primordial a escolha da metodologia a seguir neste estudo, é importante a definição da natureza da pesquisa a usar. O estudo tem por finalidade produzir conhecimentos de aplicação real na solução de problemas específicos e globais, por isso, caracteriza-se como uma pesquisa de caráter aplicada. Nesta pesquisa aplicou-se uma metodologia qualitativa, porque permitiu interlocução direta entre o utilizador e o pesquisador. “É aconselhável dar maior relevância ao aspeto subjetivo da ação social. Ao contrário da abordagem clássica ou positivista, que estabelece como regra a impessoalidade do pesquisador na investigação do fenômeno, existe um vínculo indissociável entre o mundo objetivo e a subjetividade do sujeito” (Lustosa, R. B., 2009, p.93).

A metodologia por outro lado pode ser vista como um guia de orientação seguro para o pesquisador nas suas atividades de investigação e tomada de decisões pontuais.

Flick (2004) citado em Lustosa, R. B. (2009, p. 93) afirma que “a pesquisa qualitativa não se baseia em um conceito teórico e metodológico unificado”. Ao tomar

esta citação como base de sustentação, leva a crer que vários métodos teóricos e discussão prática na pesquisa conduzem aos resultados mais precisos e com qualidade desejável. “Os métodos qualitativos consideram a comunicação do pesquisador com o campo e seus membros como parte explícita da produção de conhecimento” (Lustosa, R. B., 2009, p. 93).

O conhecimento é criado com base em observações, erros, ações, sentimentos e mais fenómenos interpretáveis. Assim, a pesquisa qualitativa aplicada, obedecendo os dois critérios “exploratória e descritiva” constitui igualmente um conhecimento. Para Verga (2004) citado em Lustosa, R. B. (2009, p.94), a “investigação exploratória deve ser realizada em área na qual haja pouco conhecimento acumulado e sistematizado”.

Nos primeiros momentos de estudo desta natureza, esta metodologia é a mais aconselhável porque permite uma melhor busca do tema em estudo, correlacionando de forma participativa, tendo em conta de que o produto se destina a um setor público, neste caso da realidade moçambicana.

A abordagem exploratória ajuda ao pesquisador a conhecer melhor o tema em estudo, permitindo-lhe maior contribuição nas propostas de soluções.

A metodologia qualitativa descritiva ajuda a descrever as características ou fenómenos do processo em estudo.

Para este caso, a primeira fase exploratória foi para melhor conhecer o ambiente do tema escolhido: modelo conceptual de Data Warehouse. A segunda fase que é a descritiva, dedicou-se a descrever o ambiente da instituição do estudo do caso. Nisto, para reforçar que o meio onde acontece a investigação científica também pode ser classificado.

De modo diferente da pesquisa quantitativa, os métodos qualitativos consideram a comunicação do pesquisador com o campo e seus membros como parte explícita da produção de conhecimento, ao invés de excluí-la ao máximo como uma variável intermédia. As subjetividades do pesquisador e daqueles que estão sendo estudados são parte do processo de pesquisa. As reflexões dos pesquisadores sobre suas ações e observações no campo, suas impressões, irritações, sentimentos, e assim por diante, tornam-se dados em si mesmos, constituindo parte da interpretação, sendo documentadas em diários de pesquisa ou em protocolos de contexto (Flick, 2004, citado em Lustosa, R. B., 2009, p. 95).

É de considerar que a metodologia qualitativa é capaz de resolver os problemas de pesquisa acadêmica, transformando-a em realidade para a utilização da sociedade.

Com esta metodologia o papel do pesquisador é participativo, não será o de relator de acontecimentos como um mero observador que não estivesse envolvido.

O pesquisador irá interagir com os envolvidos na pesquisa, seja nos seus postos de trabalho, convivência, lazer ou onde melhor convém para o participante. Como afirma Lustosa, R. B. (2009, p. 97) “Cada momento possui um objetivo específico, que ao mesmo tempo pode promover ou sofrer alteração pela reflexão da pesquisa”.

O uso de uma metodologia qualitativa descritiva num estudo conceptual de DW permite uma análise mais profunda.

Sendo o INE – M o estudo de caso, o universo da pesquisa concentrou-se numa população alvo que é a amostra dos funcionários capacitados desta instituição pública. A pesquisa no INE – M contou na sua globalidade com a participação ativa de todos os funcionários, mas na essência de recolha de dados usou-se uma pequena amostra constituída pelos gestores de topo, gestores intermédios e responsáveis seniores do departamento de informática e sistemas de informação da instituição.

Utilizou-se como instrumento de pesquisa a entrevista presencial assistida por um guião.

A tabela 3.1.1 ilustra a amostra envolvida no levantamento de dados. Considerou-se satisfatória pelo nº dos elementos da instituição envolvidos.

	Presidência	Diretores	Dir. Adjuntos	Informática
Total	3	6	5	10
Homens entrevistados	2	3	3	1
Mulheres entrevistadas		1		1
Totais entrevistados	2	4	3	2

Tabela 3.1.1 - Amostra de recolha de dados.

3.2. METODOLOGIA CONCEPTUAL DE DW

Esta pesquisa, visa encontrar uma forma melhorada de suporte à tomada de decisão, através da gestão da informação estatística produzida pelo INE – M. Para melhor atendimento ao mercado interno ou externo, a produção de informação de qualidade e atempada é indispensável numa instituição como do INE Moçambique.

Para melhor entender este tipo de base de dados é necessário tomar em consideração as metodologias dos autores considerados pioneiros deste tipo de repositório, Inmon e Kimball. A escolha de uma das metodologias de Inmon ou Kimball é apenas uma opção de arquitetura, porque nem todas as ferramentas usuais são compatíveis entre si. Por isso, cada organização deve optar pelo modelo que mais lhe convém para sua maior eficiência e desafios operacionais.

Assim, “entender os conceitos básicos da arquitetura e da metodologia de ambos os modelos proporciona um bom conhecimento acerca de *Data Warehousing*. Além de permitir adquirir conhecimentos específicos acerca das necessidades específicas de cada organização” (Carlos, F. M. M., 2008, p.62).

Inmon e Kimball são considerados os pioneiros em *Data Warehouse*, cada um com a sua metodologia. Inmon defende a metodologia top-down enquanto Kimball defende a metodologia bottom-up. A figura 3.2 ilustra as fases de qualquer uma das metodologias (Inmon ou Kimball) na conceção de DW. É importante o registo do desenvolvimento das atividades por cada fase, assim como os responsáveis de cada tarefa. Esta prática facilita a entrega do DW por fases, que vão ficando concluídas, enquanto trabalham noutras.



Figura 3.4 - Fases da Metodologia; fonte: Carlos, F. M. M. (2008, p.96)

3.2.1. Metodologia de INMON

Inmon sugere que o projeto deve iniciar com um plano de migração de dados do sistema operacional para o DW. Esta operação envolve inúmeras atividades ou tarefas, desde a limpeza de dados até ao mapeamento. O plano deve contemplar os principais assuntos ou processos, de negócio existentes na instituição. Deve também definir os relacionamentos, identificar as chaves e atributos de cada tabela. Este tipo de plano é denominado modelo de dados corporativos. “DW não deve ser construído como uma única tarefa, porque implica riscos na sua aplicação, o que pode levar a não ter sucesso no seu funcionamento. Na sua conceção deve inicialmente elaborar-se um plano interativo que permitirá o acréscimo de novos requisitos” (Inmon, 2002).

Esta metodologia define primeiro o modelo de dados corporativos, em seguida constrói-se o modelo de dados específico de cada negócio que determina os requisitos da empresa. A fase de negócios específicos é seguida pela identificação das fontes que irão alimentar de dados os modelos concebidos. “Deve escolher-se as fontes de informação, avaliando-as segundo um conjunto de critérios específicos. Como exemplos destes critérios temos: disponibilidade, qualidade, acessibilidade ou desempenho dos sistemas. Esta avaliação deverá ser realizada com rigor e análise

detalhada sobre os dados que estão disponíveis, tendo em consideração aspetos como a sua correção, redundância, acesso e a sua caracterização (metadados)” (Carlos, F. M. M., 2008).

O passo seguinte é o desenho propriamente dito do DW e a respetiva definição dos processos que terão a missão de transformar os dados de sistemas operacionais, para esta nova ferramenta. A primeira fase de construção do DW conclui-se com o povoamento deste, e com as primeiras interações com o utilizador final. Estes utilizadores devem informar todas reações que detetarem durante o uso do sistema. É através dessa informação que a equipa que desenvolveu o DW faz os reajustes e correções no sistema.

Para melhorar o funcionamento, todos os elementos que participaram na elaboração devem fazer parte da equipe de revisão. “O chefe da equipe deve ser um elemento externo ao projeto, com o intuito de dar novas perspetivas” (Inmon, 2002). A tabela 3.4.1 mostra quatro fases específicas de desenvolvimento de DW, como propõe e defende autor Inmon.

Fases	Descrição
Meth1	Dedica-se ao processo de desenvolvimento dos sistemas operacionais.
Meth2	Cuida da implementação do DW.
Meth3	Descreve o desenho e especificação dos processos de acesso ao conteúdo do DW.
Meth4	É a fase que aborda as tarefas de desenvolvimento dos processos de acesso ao conteúdo do DW.

Tabela 3.4.1 - Fases da metodologia de desenvolvimento; Fonte: adaptado de Inmon (2002)

3.2.2. Metodologia de Kimball

A metodologia de Kimball é de ciclo de vida dimensional do negócio, que consiste na criação de conceitos corporativos que possam atender todas necessidades das áreas de negócio da empresa. Para que isto seja alcançado, é exigido o comprometimento do topo organizacional, no sentido de conduzir o esforço de integração de conceitos.

A sugestão de Kimball é de escolha de um patrocinador, sendo este preferencialmente um membro da alta gerência da organização, como um diretor ou vice-presidente. Este patrocinador, por fazer uso de informações estratégicas e ser beneficiado diretamente pela aplicação OLAP, deverá conduzir o levantamento dos conceitos corporativos (Lustosa, R. B., 2009, p. 85).

Sendo este um ciclo de vida dimensional de negócios, incorpora três partes fundamentais: definição / planeamento do projeto, gestão do projeto e explicação detalhada das tarefas a realizar. A figura 3.4.2.1 mostra a interligação das três partes que formam o ciclo de vida da metodologia *bottom-up*.

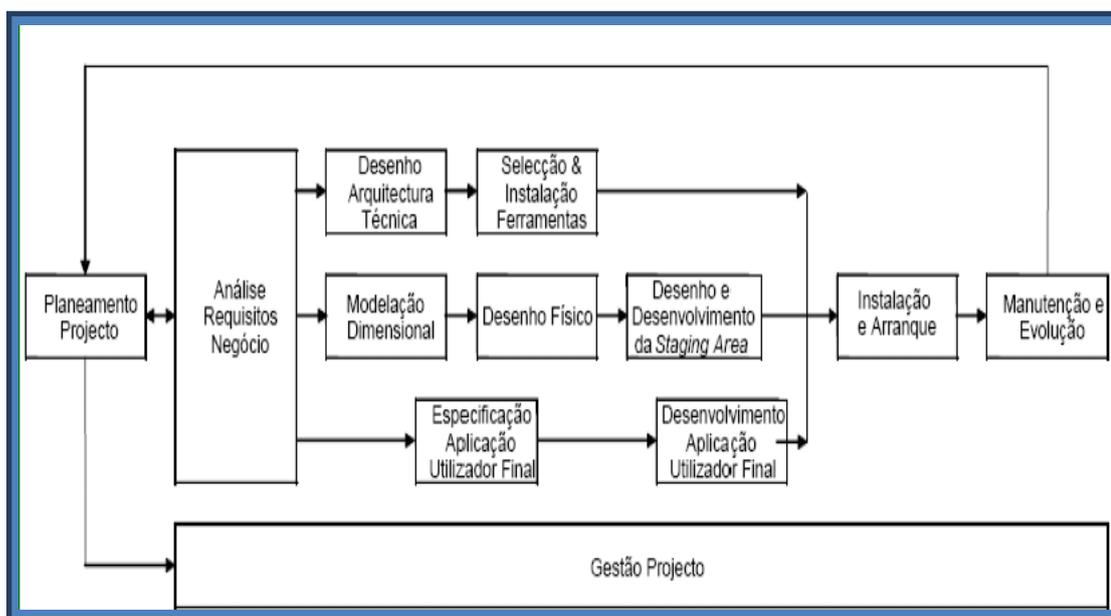


Figura 3.4.2.1 - Ciclo de vida dimensional do negócio; Adaptado de Kimball (2008)

Kimball (2002) defende uma arquitetura do tipo representado na figura 3.4.2.2 da metodologia *bottom-up*.

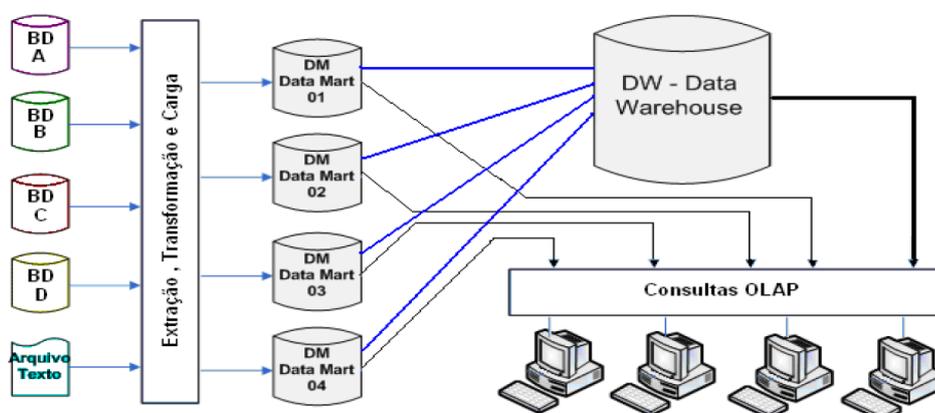


Figura 3.2.2.2 – Arquitetura bottom-up; Fonte: Adaptado de Kimball (2002)

Ao criar o DW, deve começar-se sempre pelo **plano de projeto**. É através deste plano, que se vai avaliar se a empresa ou organização está preparada para aderir a este tipo de ferramenta. Em seguida estudar a **finalidade do projeto**, em que é aconselhável a realização conjunta entre o departamento de informática e a administração, para identificar um requisito de negócio que corresponda a um só processo que tenha o maior impacto na organização.

Identifica-se igualmente um número limitado de utilizadores para esta primeira fase, os utilizadores desta fase devem ser limitados para evitar dificuldades na gestão dos requisitos. Após a fase anterior, segue-se a fase de **justificação dos requisitos**, “os parceiros devem recorrer à análise custo-benefício, assim como confrontar o retorno e o investimento, para justificar os requisitos” (Carlos, F. M. M., 2008, p. 67).

3.2.3. Comparação das Metodologias do INMON e KIMBALL

Nas duas metodologias de conceção de DW, cada uma possui os seus pontos positivos assim como de discórdia. A tabela 3.4.3 mostra os pontos de concordância e discordância entre os modelos de (Inmon e Kimball).

	Inmon	Kimball
Methodology and architecture		
Overall approach	Top-down	Bottom-up
Architectural structure	Enterprisewide (atomic) data warehouse "feeds" departmental databases	Data marts model a single business process; enterprise consistency achieved through data bus and conformed dimensions
Complexity of the method	Quite complex	Fairly simple
Comparison with established development methodologies	Derived from the spiral methodology	Four-step process; a departure from RDBMS methods
Discussion of physical design	Fairly thorough	Fairly light
Data modeling		
Data orientation	Subject- or data-driven	Process oriented
Tools	Traditional (ERDs, DISs)	Dimensional modeling; a departure from relational modeling
End-user accessibility	Low	High
Philosophy		
Primary audience	IT professionals	End users
Place in the organization	Integral part of the Corporate Information Factory (CIF)	Transformer and retainer of operational data
Objective	Deliver a sound technical solution based on proven database methods and technologies	Deliver a solution that makes it easy for end users to directly query the data and still get reasonable response times

Tabela 3.4.3 – Comparação das Metodologias de Inmon e Kimball; Fonte: Breslin, M. (2004)

3.3. REQUISITOS DO DATA WAREHOUSE

Teoricamente não é aconselhável no levantamento de requisitos optar por um único método, porque as duas metodologias vistas anteriormente têm vantagens assim como possuem igualmente pontos de discórdias. É necessário ter em mente as necessidades dos diferentes utilizadores e das características dos sistemas atualmente a operar na organização. A informação dos requisitos nesta fase pode ser funcional ou não funcional, mas deve estar na tabela de análise de requisitos.

Kimball sugere a identificação, separação e classificação dos requisitos funcionais e não funcionais. Sugere igualmente que seja feita de forma conjunta entre os técnicos de informática e os utilizadores, para melhor identificar as responsabilidades. “Os critérios de identificação de requisitos definidos nesta fase permitem capturar atributos para cada tipo de requisito, a partir dos quais é possível gerir a inclusão desses requisitos no sistema” (Carlos, F. M. M., 2008, p.78).

Os principais pontos que devem constar no plano geral de requisitos são: Projeto; Funções/Responsabilidades; Procedimentos; Requisitos funcionais e não funcionais.

Projeto – identifica o projeto em estudo

Funções e responsabilidades – identifica a função e a responsabilidade de cada indivíduo que faça parte da gestão dos requisitos do projeto.

Procedimentos – descreve as regras de como será a gestão dos requisitos do sistema, tais como, mudanças de conteúdo; testes dos requisitos; validações, assim como especificações de novos requisitos.

Requisitos (Funcionais e não Funcionais) – para funcionais, a equipa deve enumerar todos levantamentos efectuados, enquanto não funcionais, deve existir uma lista predefinida. “Ambos os requisitos devem ser classificados e em seguida deverá ser elaborada uma lista de regras para a seleção dos mesmos” (Kimball, 2008).

4. INSTITUTO NACIONAL DE ESTATÍSTICA DE MOÇAMBIQUE

Este capítulo tem como teor a descrição sumariada do que é o Instituto nacional de Estatística de Moçambique (INE – M), sua missão, visão, valores, desafios e necessidades de informação como produtora da informação estatística oficial do País.

4.1. O INE – M COMO ORGANIZAÇÃO

O Instituto Nacional de Estatística de Moçambique foi criado pelo Decreto Presidencial, de 28 de Agosto de 1996. “É uma instituição coletiva de direito público, dotada de personalidade jurídica, autonomia técnica, administrativa e financeira, tutelado pelo Conselho de Ministros”.

Independentemente das esferas de atividade particulares, a maior parte das organizações apresenta estruturas de organização semelhantes, cuja estrutura típica apresenta-se normalmente em forma de triângulo (Julião, 2009). Segundo o mesmo autor, na base do triângulo encontra-se o nível Operacional onde se processa a produção.

O nível intermédio constitui-se pelos gestores, investigadores e administradores cujas tarefas incluem a monitorização do desempenho do nível da base (Operacional), a pesquisa do ambiente externo à organização e a preparação de opções de política para o nível de topo.

O nível executivo constitui o topo do triângulo onde são determinadas, através de um pequeno grupo de decisores, as orientações estratégicas da organização.

Segundo Mintzberg, referido em Julião (2009), o triângulo organizacional é, normalmente, referido em organizações industriais. Contudo, o triângulo pode ser facilmente traduzido para considerar serviços e organizações públicas onde produtos e serviços são fornecidos aos clientes e cidadãos.

De facto, o INE-Moçambique é uma organização pública e apresenta estrutura organizacional semelhante ao observado pelo autor. É composto por três níveis nomeadamente, presidência, direções dos serviços centrais e departamentos dos serviços centrais (figura 4.1.1 e 4.1.2).

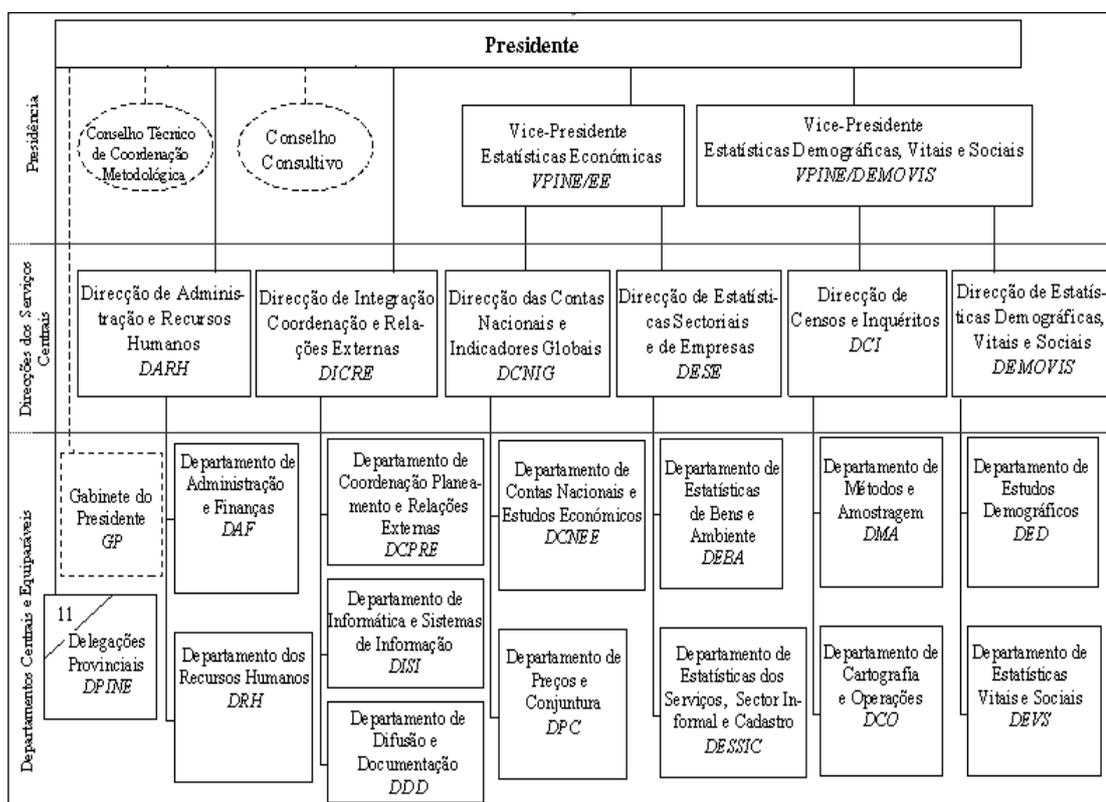


Figura 4.1.1 - Organograma do INE – M; fonte: INE – Moçambique

Ao nível do topo, o INE – M é constituído pela presidência que é um órgão consultivo do presidente para a gestão estratégica e operacional da instituição. Ela é composta pelo Presidente que o dirige, e por dois vice-presidentes nomeadamente, o Vice-Presidente para o Pelouro Económico e o Vice-Presidente para o Pelouro Demográfico.

As direcções, num total de seis, constituem a parte intermédia do triângulo cujas funções são as de planeamento, coordenação das relações inter e intrainstitucionais, gestão, difusão e marketing de informação, análise e investigação demográfica e social, provisão de serviços gerais de administração, gestão e desenvolvimento dos recursos humanos e formação de quadros.

Os departamentos dos serviços centrais são a base do triângulo (nível operacional) onde são exercidas todas as tarefas relacionadas com a produção das estatísticas.



Figura 4.1.2 - Estrutura organizacional do INE; adaptado de Mintzberg em Julho (2009)

4.2. MISSÃO DO INE – M

A missão do INE segundo João Dias Loureiro (Presidente do INE-M) é a de responder em tempo oportuno às necessidades de informação dos utilizadores e promover a cultura estatística nacional.

4.3. VISÃO DO INE – M

Ser a principal referência estatística para o desenvolvimento nacional, segundo o Presidente da instituição.

4.4. VALORES DO INE – M

Para esta instituição de aparelho do estado, o seu valor acrescentado está direcionado aos utilizadores, o profissionalismo e as sinergias.

4.5. DESAFIOS DO INE – M

O INE de Moçambique, segundo o seu dirigente máximo, João Dias Loureiro, tem como desafio principal os seguintes pontos:

Monitoria do Plano de Ação para a Redução da Pobreza Absoluta (PARPA) e dos planos do Governo

Exigência de mensuração da evolução dos níveis da pobreza, da economia, da produtividade que implicam uma vasta utilização de estatísticas multidisciplinares, para as quais o Sistema Nacional de Estatística (SEN) deve desempenhar um papel ativo. Deve-se ainda aperfeiçoar o quadro do planeamento estratégico, tático e operacional do SEN e seu alinhamento com o quadro geral do sistema de planeamento público.

Descentralização

Fortalecer a capacidade de produção estatística ao nível das províncias e dos distritos e municípios que respondam a necessidades do processo de governação e desenvolvimento local e nacional.

Consolidar a produção de estatística prioritária

Racionalizar as funções dos órgãos produtores de estatísticas para responder aos objetivos do PARPA II, melhorar a coordenação intersectorial, e evitar redundâncias. Assegurar que os processos de reforma das instituições da função pública ao nível central e local acautelem devidamente a função estatística de modo a que deles resultem designadamente, registos administrativos que possam servir de fontes de informação para o Sistema Estatístico Nacional.

Parcerias

Formar parcerias de modo a assegurar um financiamento apropriado para as atividades estatísticas face a um eventual cenário restritivo do Orçamento do Estado, tirando vantagens da abertura a fluxos adicionais de ajuda externa, desde que alinhados com as atividades e planos do Governo.

Fortalecer os Recursos Humanos

Criar um quadro atraente de desenvolvimento, gestão e retenção dos seus quadros designadamente da massa crítica. Desenhar ações de formação para o nível central e nível local para garantir as capacidades necessárias para a implementação do Plano Estratégico.

Novas tecnologias

Tirar vantagens dos novos desenvolvimentos das tecnologias de informação e comunicação, estabelecendo um quadro apropriado no qual estas impulsionem a produção, troca e disseminação das estatísticas de forma integrada e atempada nos níveis multi-institucionais, nacional, provincial e distrital. No plano nacional, haverá que considerar as oportunidades oferecidas pelo GovNet.

Revisão do quadro legal (lei do Sistema Estatístico Nacional)

No quadro geral do desenvolvimento socioeconómico do País a atual lei do SEN mostra-se inadequada para responder aos desafios que se apresentam no âmbito da necessidade da informação estatística. Pelo facto urge a necessidade da revisão do quadro legal que é a lei do Sistema Estatístico Nacional.

A construção de um edifício próprio

O presidente do INE assim como os funcionários contactados sentem grande alívio e satisfação por este ponto ter sido satisfeito desde Agosto de 2011. A qualidade da informação estatística tem muito a ver com o meio e condições na sua produção.

O melhoramento das Estatísticas Económicas

Especificamente as relativas a dados empresariais, estabelecimento duma base de dados do SEN compreensiva e consolidada, e desenvolvimento de uma estratégia de comunicação e disseminação de informação estatística mais efetiva.

4.6. NECESSIDADE DE INFORMAÇÃO

Por falta de uma metodologia uniformizada de armazenamento e gestão de dados na instituição, tem sido difícil o nível operacional responder com eficiência e eficácia aos intermediários, assim como os intermediários ao topo. Sendo o topo o nível decisor, este tem enfrentado imensas dificuldades em tomar uma decisão segura, confiante e rápida.

Segundo João Loureiro, com uma boa informação, segura e eficiente pode-se:

Responder às necessidades estatísticas dos utilizadores sobre a estrutura e tendência da população, economia, sociedade, governação, género e ambiente e promoção na utilização da informação estatística oficial.

Criar progressivamente uma capacidade de produção estatística ao nível dos distritos e municípios que respondam às necessidades do processo de governação e desenvolvimento local e nacional.

Capitalizar a produção das estatísticas com recurso a novas tecnologias e metodologias, inovação, coordenação interinstitucional, apropriação de registos administrativos e redução da carga de respostas da origem de dados.

Aprimorar a capacidade institucional face aos novos desafios impostos pelos utilizadores.

Face a estes pontos, fica claro que DW irá ajudar de forma significativa na supressão de alguns pontos focados pelo presidente do Instituto Nacional de Estatística de Moçambique (João Dias Loureiro), e não só, o DW permitirá conservar dados históricos que posteriormente poderão ser usados para vários fins.

4.7. GESTÃO ESTRATÉGICA

O INE-M é uma instituição de aparelho do estado, e para o seu funcionamento normal depende de vários fatores (humano, financeiro, parcerias e tecnológico). Para uma melhor gestão estratégica, é de conciliar através de uma análise SWOT, de forma a identificar os pontos fortes, oportunidades, fraquezas e ameaças.

Forças <ul style="list-style-type: none">- Grande produção de informação estatística de qualidade.- Partilha com eficácia e coesão da informação produzida.- Boa eficiência nos serviços prestados.- Acesso rápido e seguro de dados de apoio à tomada de decisão.	Fraquezas <ul style="list-style-type: none">- Muitos dados de produção estatística dispersos.- Quadros ou técnicos qualificados ainda em número muito reduzido.- Infraestrutura tecnológica ainda muito reduzida.
Oportunidades <ul style="list-style-type: none">- Integração e normalização de dados de diversas origens.- Manutenção de parcerias internas e externas.- Formação de equipas de trabalho com espírito de partilha de conhecimentos.- Boa gestão do espaço físico.- Os diversos processos administrativos orientados a processos de negócios da empresa.	Ameaças <ul style="list-style-type: none">- Incapacidade financeira para acompanhar a evolução tecnológica.- Incompetência técnica de alguns quadros qualificados pode comprometer a credibilidade e o crescimento do INE.- Mudanças socioeconómicas muito rápidas não previstas.- Falta de adequada manutenção.

Tabela 4.7.1 - Análise SWOT

Sendo o INE uma instituição do aparelho do estado com a finalidade de atender prontamente todas as camadas, socioeconómicas, políticas e estudantis, deve procurar sempre manter a informação constantemente atualizada e disponível. Por isso, é

fundamental possuir um suporte que possibilite um planeamento, comunicação e gestão de fatores estratégicos. A figura 4.7.1 mostra o resumo das perspetivas estratégicas, que resultam em sucesso quando são bem implementadas na organização.

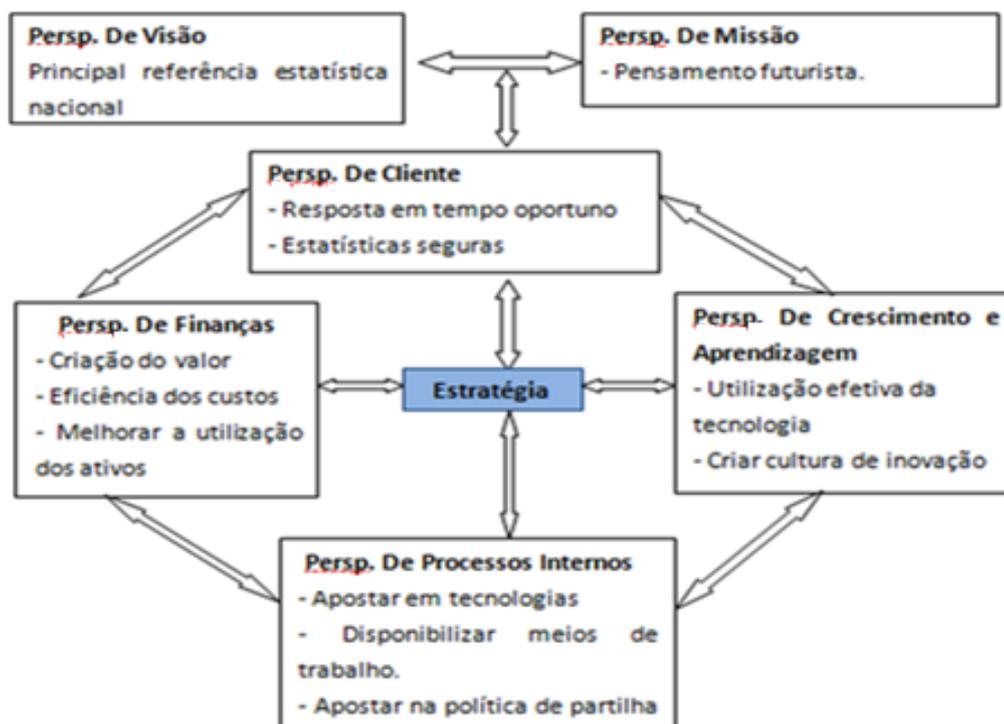


Figura 4.7.1 - Perspetivas estratégicas; Adaptado de Silva, C. I. M. (2011)

4.8. REQUISITOS DO PROJETO

Os requisitos deste projeto irão ajudar a ter uma visão abrangente das atividades do INE-M ao longo do tempo, dos indicadores e métricas que revelam o comportamento e desempenho dos negócios. “Um sistema de apoio à decisão deve refletir os objetivos e metas da organização através de visão dos seus tomadores de decisão” (Ziulkoski, L.C.C., 2003, p. 11). Alguns autores consideram que os requisitos deveriam ser levantados depois de o DW ser parcialmente carregado e usado pelos decisores da empresa. Porque assim, poderiam informar nomeadamente, sobre os dados desnecessários na base e contribuir para outros desenvolvimentos do DW, o que ajudaria no refinamento e controlo de riscos do projeto.

A teoria anterior aponta sobretudo para o facto de ser arriscado iniciar um projeto de DW sem considerar primeiro os requisitos dos utilizadores. Os autores posteriores a Inmon são de opinião que devem coletar-se requisitos antes de implementar DW numa organização. Neste estudo a coleta de requisitos foi feita antes, usando como ferramenta de recolha a entrevista auxiliada por um guião, como se referiu no item 3.1, do capítulo da metodologia.

Dos 11 entrevistados, quanto ao seu nível de escolaridade: 2 bacharelato, 4 licenciatura e 5 com o mestrado. Tempo de serviço como funcionários na instituição: 1 está há menos de 5 anos, 3 entre 10 a 15 anos e os restantes 7 contam com mais de 20 anos de serviço. Experiência como gestor: 1 menos de três anos, 6 entre 5 a 10 anos e 4 já contam com 10 ou mais anos como chefes em diversos setores da instituição.

Para melhor se inteirar dos requisitos, as entrevistas foram acompanhadas por um guião e divididas em 3 grupos: Gestores do topo, gestores intermédios e departamento de informática. Isto porque cada nível de gestão possui uma visão diferente apesar de a finalidade ser a mesma, o crescimento da instituição.

Nos três níveis de entrevistas teve-se como resultados:

- O nível estratégico considera como objetivo principal responder pontualmente e com eficácia aos utilizadores da informação estatística. Para os intermédios é de responder de segura e atempada ao nível estratégico do topo, para que estes façam uso da informação.
- Consideram ser comum notar o não cumprimento das tarefas previstas, pela dificuldade que surge quando se trata de prestar contas aos superiores hierárquicos, ou quando se pretende responder um pedido formulado pelos clientes. Isto tem sido mais frequente quando se trata de informação sobre as estatísticas económicas empresariais.
- Todos os entrevistados consideram não existir ferramenta de apoio à tomada de decisão na instituição.
- Quanto à recolha de dados, afirmam ser por questionário, mas há ocasiões em que os dados são enviados em formato Excel.

- Os entrevistados foram unânimes em afirmar que os dados são recolhidos e tratados mensalmente.
- A infraestrutura do INE – M é composta por 6 servidores: dados, web, correio eletrônico, intranet, publicações e servidor de aplicações. Este último ainda não povoado. A rede de computadores existente no INE-M é local e por cabo. Conta ainda com cerca de 200 computadores e impressoras multiusos. Usufrii da rede do governo (govNet), mas esta, é limitada para alguns assuntos governamentais.
- A maior parte dos softwares usados na instituição são da tecnologia Microsoft, tais como: Windows server 2008, Windows7, Windows XP, Microsoft Office, Microsoft Visual Studio, Microsoft Visio, Microsoft Share Point e Microsoft SQL Server 2008, licenças doadas na sua maioria pelos parceiros da instituição.
- Quanto a aplicações operacionais, verificou-se existência de um número considerável e de forma desorganizada, algumas concebidas na instituição e outras doadas pelos parceiros. Em todas as aplicações, nota-se a ausência de documentação técnica.
- Quanto à implementação de DW, todos se afirmaram favoravelmente, prontificando-se em dar todo apoio necessário para que o projeto tenha sucesso.

5. MODELAÇÃO DO DATA WAREHOUSE PARA O INE-M

Esta é a fase da descrição de todos os conceitos necessários para conceber o DW proposto. “O modelo conceptual descreve os conceitos, em diferentes perspetivas de análise e como os conceitos se relacionam entre si. O modelo lógico – descreve a forma de organização da informação no data warehouse: Molap, Rolap, etc... é já da responsabilidade do fornecedor do SGBD” (Quaresma, S., 2004).

A documentação do projeto constitui um elemento chave para monitorizar e garantir uma boa gestão de projeto. É difícil negociar um projeto se não existir uma documentação que caracteriza e regista todas as decisões, problemas e soluções. Na prática são utilizadas duas técnicas: através de reuniões periódicas com a equipa ou através de relatórios periódicos informando a equipa e os patrocinadores do decorrer do projeto.

Na construção deste modelo é importante a verificação dos prazos e tarefas por parte dos responsáveis dos SI. Como forma de redução de riscos na construção de DW, consideram três fatores: Comunicação, Planeamento e Qualidade de dados.

5.1. MODELO PROPOSTO

Após análise e verificação da revisão da literatura, dos requisitos, e das metodologias de Inmon e de Kimball, propôs-se para o estudo de caso, o modelo de DW em estrela. O desenvolvimento deste modelo pode ser subdividido em várias atividades, algumas desenvolvidas em paralelo, outras de forma sequencial, e outras de forma contínua. A forma contínua requer acompanhamento durante o decorrer de todo projeto do DW. Este modelo foi concebido por Ralph Kimball, usando a metodologia *bottom-up* e é baseado em *Data Marts* departamentais ou por assunto.

De acordo com os levantamentos realizados, seria difícil ao INE-M adotar o modelo *top-down* de Inmon, por vários motivos, desde o aspeto técnico ao financeiro. O modelo *bottom-up* sugerido permite a construção com os escassos recursos que a instituição possui. Isto é possível porque o modelo para além de simples na sua implementação, permite uma construção incremental de *Data Marts*.

A figura 5.1.1 ilustra o modelo da arquitetura sugerido ao INE-M.

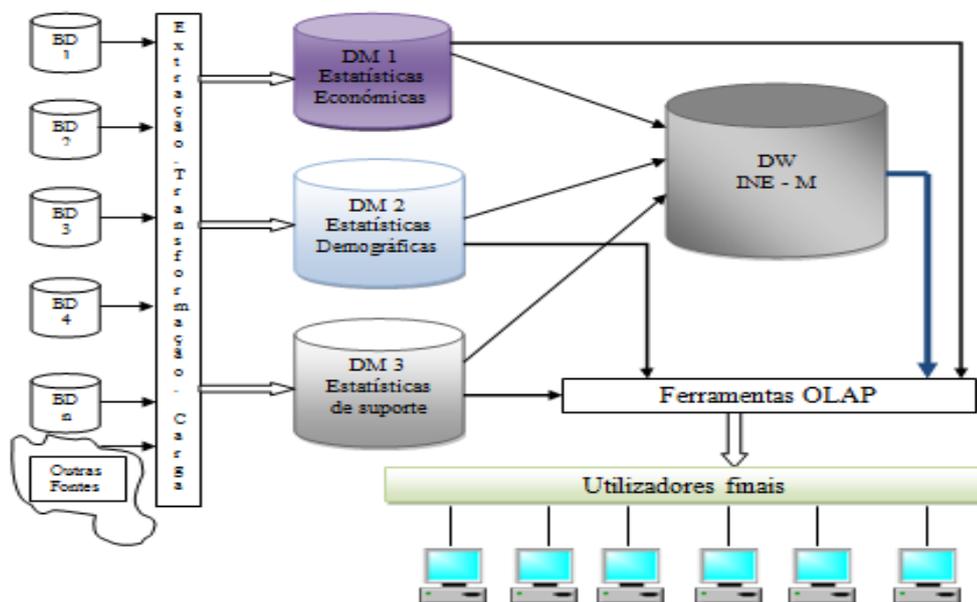


Figura 5.1.1 - DW para INE-M; Fonte: Adaptado de Kimball (2002).

O modelo em estrela, mais acessível para o utilizador final, vai crescendo há medida que os interesses de negócios da instituição crescem, e é de fácil manutenção.

O DW do INE será constituído por três DMs departamentais ou por assuntos que são: Estatísticas económicas, Estatísticas demográficas e Estatísticas de suporte. Os *Data Marts* serão concebidos de forma *bottom-up* como mostra a figura 5.1.1, com possibilidade de acréscimos futuros se os processos da instituição assim exigirem.

Na primeira fase será modelado e implementado um DM, e na medida que vai sendo usado serão registadas todas as irregularidades que surgirem, assim como, serão registados os metadados não identificados anteriormente. Isto permitirá que a equipa envolvida na revisão corrija esses problemas de modo a não se refletirem nos DMs posteriores. Na base nos requisitos, sugeriu-se o DM das estatísticas económicas como o primeiro a ser modelado.

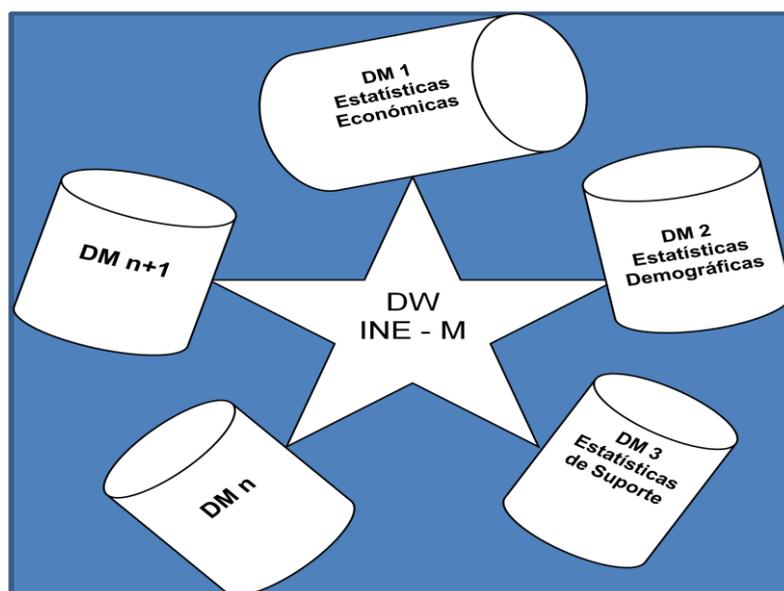


Figura. 5.1.2 - Modelo dimensional do DW em estrela para o INE-M

Antes de iniciar o processo de modelagem do DM, vai-se desenhar um modelo representativo entidade relacionamento (E-R), de duas aplicações operacionais em uso na instituição. As aplicações operacionais e dados externos existentes no INE-M servirão de suporte básico para implementação do DW. Vai efetuar-se essa modelagem concetual de pelo menos duas aplicações, por não existir nenhuma documentação técnica dessas aplicações na instituição.

5.2. MODELO E-R DE SISTEMAS OPERACIONAIS

Por as aplicações existentes não possuem documentação, não é fácil compreender o funcionamento das mesmas. De todas as aplicações das estatísticas económicas empresariais existentes, seleccionaram-se aleatoriamente duas aplicações. Essas serão representadas e modeladas neste trabalho no formato relacional.

A figura 5.2.1 representa um conjunto de entidades e seus atributos. “Uma entidade é algo do mundo real que possui uma existência independente. Uma entidade pode ser um objeto com uma existência física - uma pessoa, carro ou empregado - ou pode ser um objeto com existência conceptual - uma companhia, um trabalho ou um curso universitário. Cada entidade tem propriedades particulares, chamadas atributos” (Takai, O. K.; Italiano, I. C.; Ferreira, J.E., 2005, p. 23).

Tab_Prod_Industrial			Tab_Prov			Tab_Empresa		
Nome	Tipo	Tamanho	Nome	Tipo	Tamanho	Nome	Tipo	Tamanho
Mês	Número	2	Id_prov	texto	2	Id_empresa	número	6
Ano	Número	4	Prov	texto	15	Nome_Emp	texto	100
Id_Empresa	Número	6	Tab_Distritos			Id_TipoEmpresa	número	1
CodCAE	Texto	5	Nome	Tipo	Tamanho	Id_Local	texto	2
Id_Prod	Texto	8	Id_Distrito	texto	2	Ano_Constituicao	número	4
NPS_Nac_H	Número	5	Id_Prov	texto	2	Id_CAE	texto	5
NPS_Nac_M	Número	5	Distrito	texto	20	Id_FJR	número	1
NPS_Estrang_H	Número	5	Tab_Postos			NUIT	número	9
NPS_Estrang_M	Número	5	Nome	Tipo	Tamanho	Tab_ComercioServicos		
Remun_Nac_H	Moeda	10	Id_Posto	texto	2	Nome	Tipo	Tamanho
Remun_Nac_M	Moeda	10	Id_Distrito	texto	2	Mês	Número	2
Remun_Estrang_H	Moeda	10	PostoAdm	texto	50	Ano	Número	4
Remun_Estrang_M	Moeda	10	Tab_Produto			Id_Empresa	Número	6
Quant_prod	Número	5	Nome	Tipo	Tamanho	CodCAE	Texto	5
Preco_Medio	Moeda	10	Id_prod	texto	8	Id_Prod	Texto	8
Valor_Producao	Moeda	10	Produto	texto	50	NPS_Nac_H	Número	5
Valor_Venda	Moeda	10	Id_tipoProd	número	1	NPS_Nac_M	Número	5
Tab_Localidades			Id_medida	número	1	NPS_Estrang_H	Número	5
Nome	Tipo	Tamanho	Id_ProdOrigem	número	3	NPS_Estrang_M	Número	5
Id_Local	texto	2	Tab_CAE			Remun_Nac_H	Moeda	10
Id_Posto	texto	2	Nome	Tipo	Tamanho	Remun_Nac_M	Moeda	10
Localidade	texto	50	Id_CAE	texto	5	Remun_Estrang_H	Moeda	10
Tab_TipoProduto			CAE	texto	100	Remun_Estrang_M	Moeda	10
Nome	Tipo	Tamanho	Tab_UnMedida			Encardo_Social_Obrig	Moeda	10
Id_TipoProd	número	1	Nome	Tipo	Tamanho	Encardo_Social_Faculta	Moeda	10
Tipo_produto	texto	50	Id_Medida	número	1	Encardo_Formacao	Moeda	10
			Medida	texto	15			

Figura 5.2.1 - Atributos e caracterização de cada entidade

Caracterizar atributo entende-se por criar tabelas ainda na fase inicial, onde se identificam os possíveis candidatos a atributos das entidades. A figura 5.2.1 mostra o exemplo de caracterização em forma de tabelas. As tabelas representam as entidades de base de dados e os campos representam atributos. Os atributos pintados representam nesta fase as possíveis chaves das entidades. As figuras (5.2.2 e 5,2,3) a seguir representam E-R e cardinalidade entre as entidades.

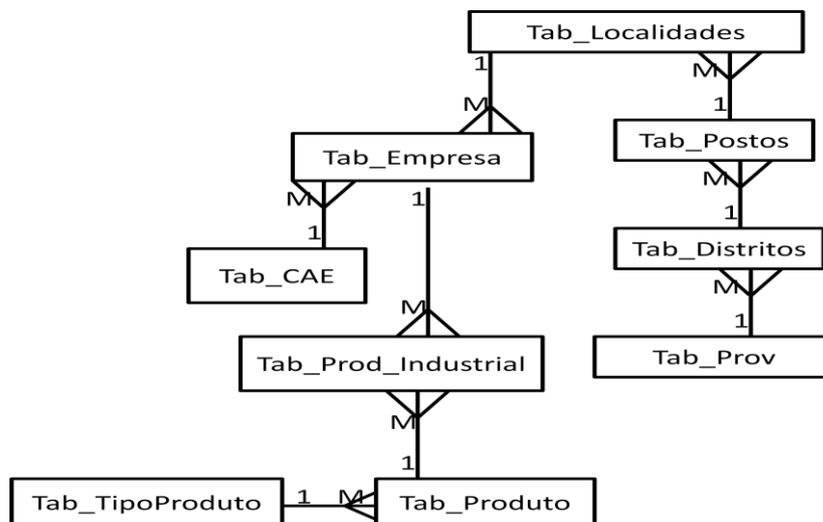


Figura 5.2.2 - E-R, cardinalidade e hierarquia da produção Industrial

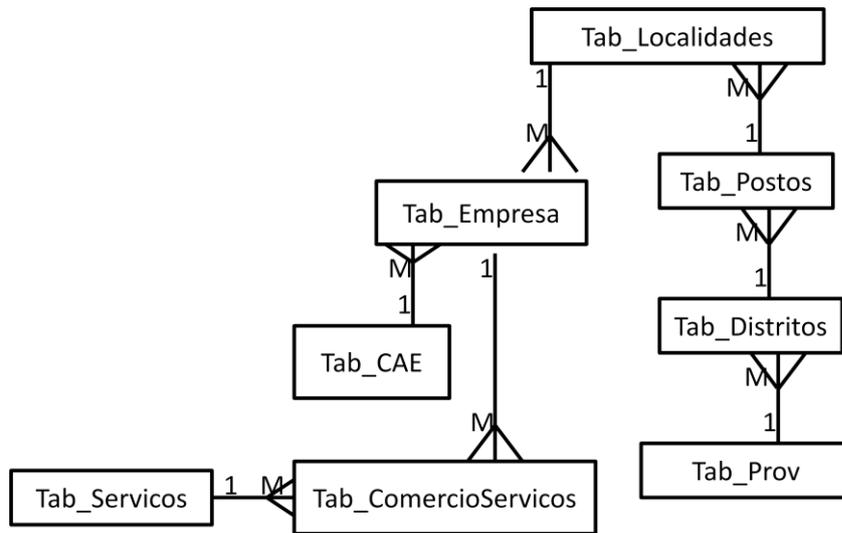


Figura 5.2.3 - E-R, cardinalidade e hierarquia de comércio e serviços

A relação entre entidades é a fase de modelagem de dados, onde se procura fazer a ligação entre entidades segundo o seu relacionamento, como mostra a figura 5.2.2 e 5.2.3. Nestas imagens os retângulos representam as entidades e as linhas representam o elo de ligação entre uma entidade e outra. A extremidade da linha terminada por três traços significa muitos e onde termina com apenas um traço significa um. Correspondendo assim uma relação com cardinalidade de “um-para-muitos”. As figuras 5.2.4 e 5.2.5 e a representação E-R desenhados em Access.

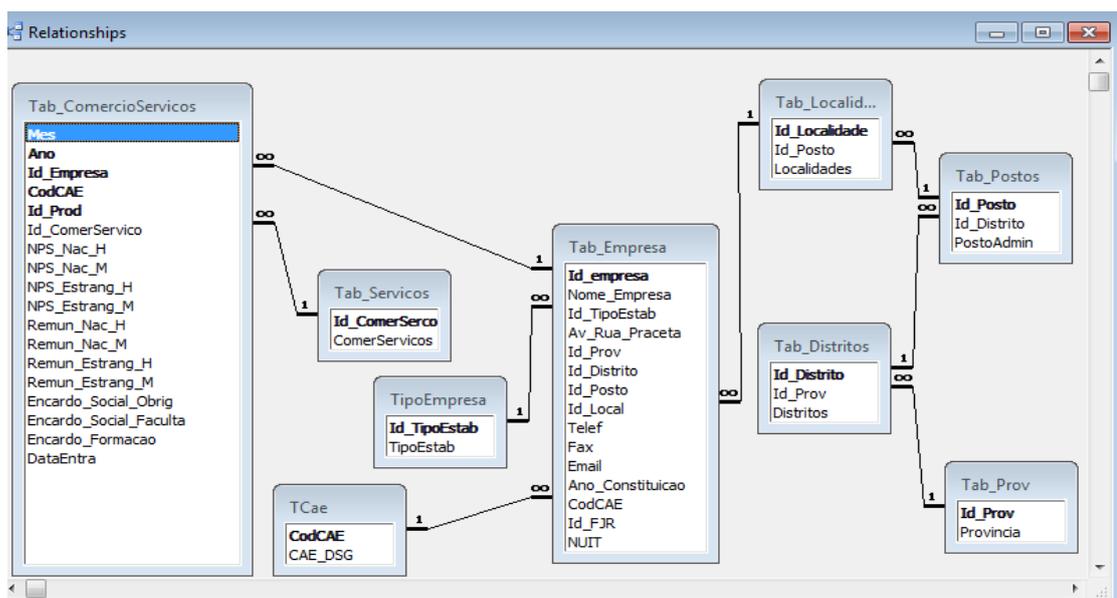


Figura 5.2.4 – Diagrama E – R Comércio e Serviços

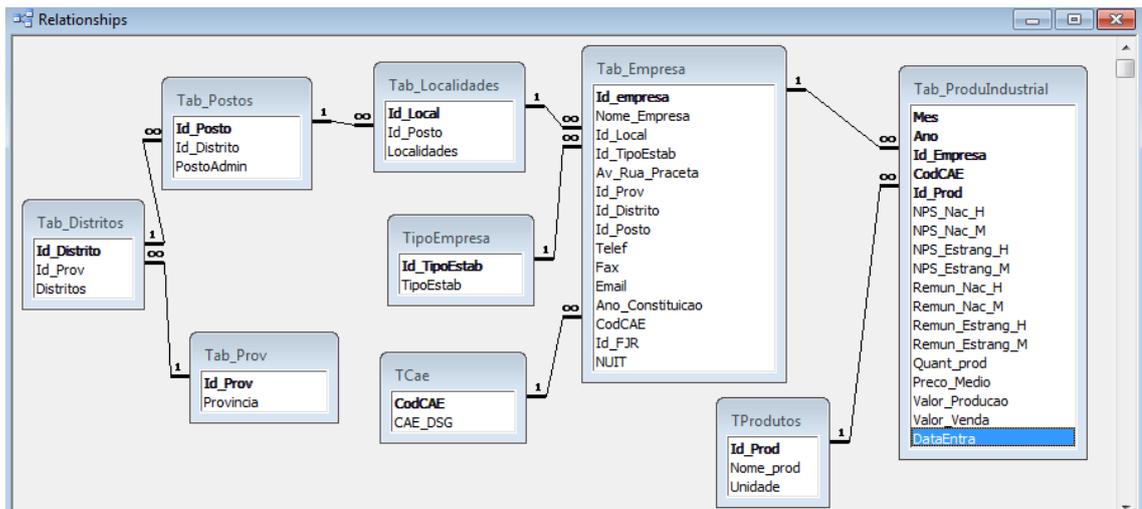


Figura 5.2.5 – Diagrama E – R de estatísticas de Produção Industrial

Com os dados das aplicações operacionais e dos dados externos, faz-se o carregamento, transformação e busca do DM ou DW.

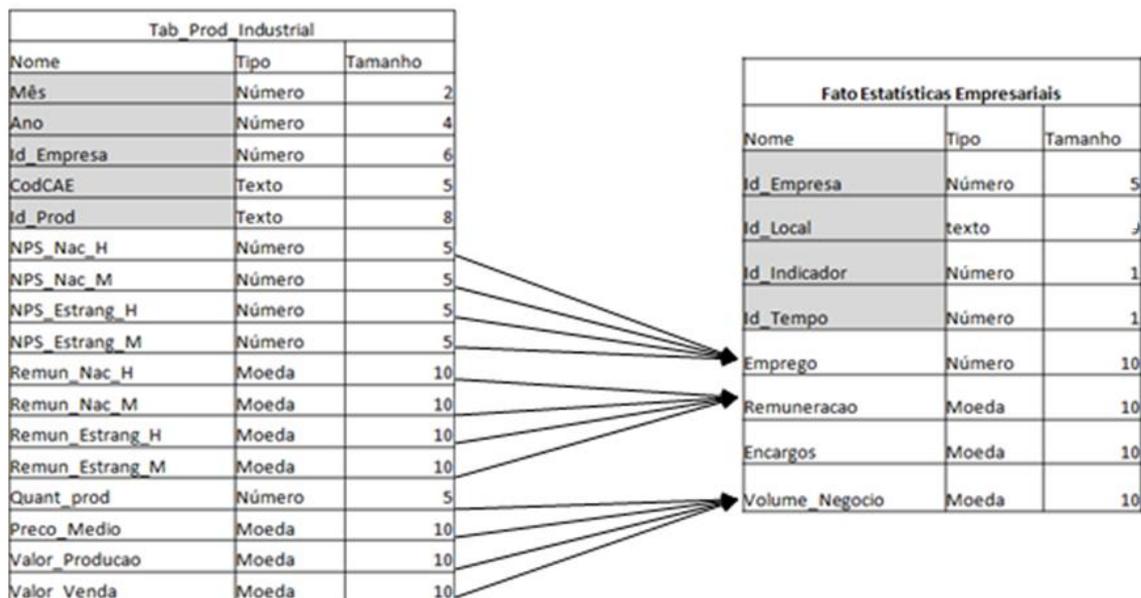


Figura 5.2.6 - Exemplo de transformação e aglomeração de dados.

5.3. MODELO DIMENSIONAL DE DW

Este modelo obedece a uma metodologia de 3 etapas nomeadamente:

1. Classificação de entidades.
2. Identificação de hierarquias.
3. Projeção dimensional.

Classificação de entidades – A entidade classifica-se em 3 categorias nomeadamente “Transação, Componente e Classificação”. É necessário observar que cada modelo E-R deverá estar contido numa das 3 categorias de entidades (de Transação, de Componente e de Classificação).

A entidade de transação é aquela que contém o registo de acontecimentos de negócios da empresa realizadas no tempo e que podem ser sumarizadas. A entidade componente detalha ou especifica cada negócio registado na entidade de transação, tendo uma relação na entidade de relacionamentos de “um-para-muitos”, responde perguntas como por exemplo, “Quando ...? Quem ...? O que ...? Onde ...?”.

Entidades de classificação é diretamente ligada a entidade componente, com relacionamento de “um-para-muitos”, são as usadas na formação de tabelas dimensão no modelo estrela. A figura 5.3.1 mostra a classificação das entidades usadas no modelo de dados “no diagrama as entidades em preto representam entidades de Transação. As entidades em cinza representam entidades Componente e as entidades em branco representam entidades de Classificação” (Dill, 2002, p.49).

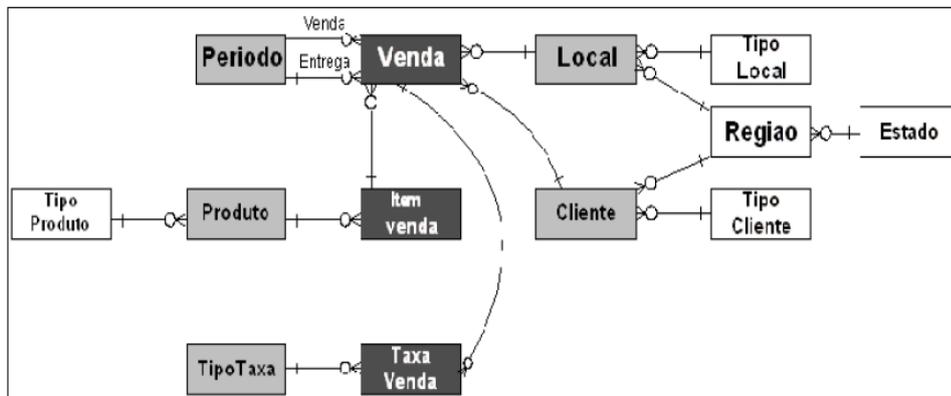


Figura 5.3.1 Categorias de entidades; Fonte: Dill (2002), adaptado de Moody & Kortink, (2000)

Identificação hierárquica – é a derivação dos modelos E-R em dimensional, pode-se afirmar que é o sequenciamento de ligações das entidades obedecendo uma relação de “um-para-muitos” com a mesma direção. A figura 5.3.2 mostra o exemplo de hierarquia. Na classificação de entidades, a de transação encontra-se hierarquicamente a nível mais alto, enquanto a de classificação está a nível mais baixo.

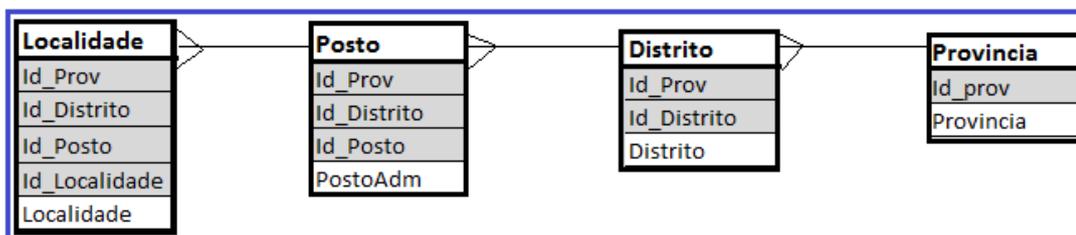


Figura 5.3.2 - Hierarquia de um modelo conceitual

Projeção dimensional – é a fase da conceção da estrutura do DW, nesta fase ainda identifica-se os atributos e chaves das tabelas de facto e dimensões. Como o projeto proposto é o modelo *bottom-up*, neste estudo vai-se fazer a modelagem apenas de um Data Mart das Estatísticas económicas empresariais como se afirmou acima, por ser mais representativo e que constitui a maior preocupação até este momento dos gestores do INE-M.

6. DATA MART

6.1. CARATERIZAÇÃO

Considera-se DM uma fragmentação de DW destinado a responder um determinado assunto da empresa. Aliás pode-se dar a mesma definição de DW aos DMs com a ligeira diferença de o DW estar indicado para os assuntos de negócio de toda organização, enquanto o DM é voltado para uma determinada área ou assunto específico da empresa.

As principais características das duas bases são o fácil acesso da informação, consistência, flexibilidade para efetuar alterações adaptadas às mudanças, segurança da informação, e uso da informação para apoio à tomada de decisão.

Cada DM é composto de tabelas dimensão e de facto assim como o DW, é de referir que a tabela dimensional pode possuir apenas uma ligação com a de facto, e a de facto é única e liga-se com diversas tabelas de dimensão no DM. Geralmente as tabelas de dimensão são caracterizadas de muita redundância por não estarem de forma normalizada. Enquanto a de facto não contém redundância, é fortemente normalizada.

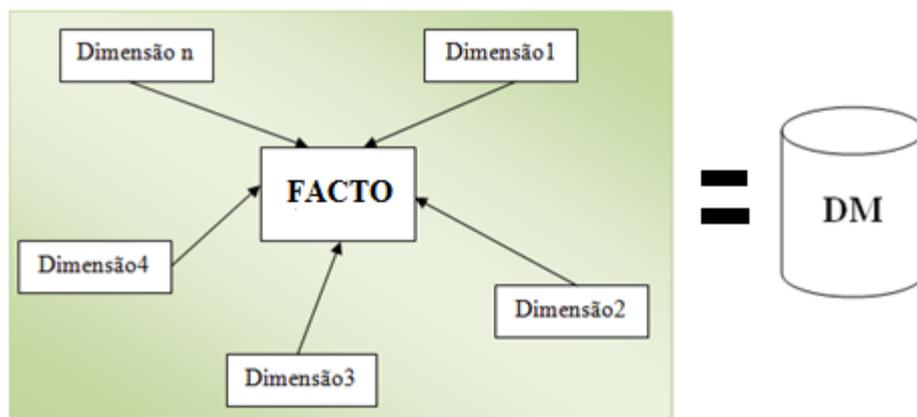


Figura 6.1.1 - Constituição de DM com Tabelas de facto e dimensionais

6.2. CRIAÇÃO DE DM

O processo de criação de *Data Mart* é constituído por quatro passos essenciais: Seleção do processo de negócio; definição da granularidade; escolha das dimensões e identificação dos factos (Kimball, 2002).

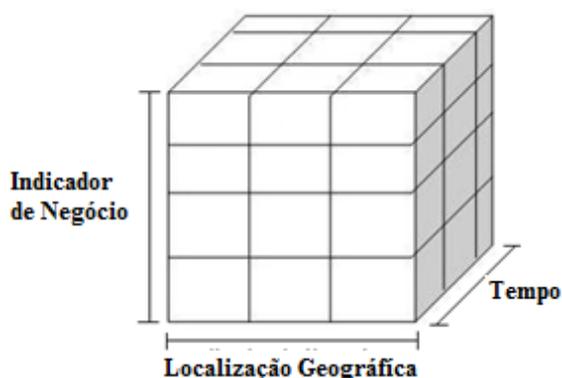
6.2.1. Seleção do processo de Negócio

Dos três Data Marts que irão constituir o DW institucional, a seleção do processo de negócio, recaiu para as estatísticas económicas de empresas, por ser a maior preocupação dos gestores da instituição. As dificuldades no controlo dos dados desta área têm sido muito complicada. Os gestores intermédios desta área dificilmente respondem pontualmente com os dados existentes aos seus superiores hierárquicos. A conceção dos relatórios tem levado muito tempo devido às dificuldades de conciliação e uniformização de dados.

Com o uso de DM neste pelouro estatístico, irá minimizar de forma significativa seja na qualidade dos dados, seja no tempo de resposta dos relatórios de apoio à tomada de decisão.

6.2.2. Definição da Granularidade

O nível de granularidade é importante no DW por ser a medida determinante da qualidade dos dados, e afeta a velocidade e o armazenamento. Neste DM definiu-se como nível de granularidade o mês, ou seja os dados agrupam-se de janeiro a dezembro. Este agrupamento é conjugado por intermédio das tabelas de dimensão Localização geográfica, Indicador do Negócio e Tempo. Decidiu-se assim, por ser aconselhável o uso dentro do possível os níveis mínimos de granularidade desde que não use espaços maiores de armazenamento.



6.2.3. Escolha das Dimensões

Com os levantamentos efetuados na instituição, considerou-se como dimensão deste DM, quatro tabelas que são: Dim_Indicadores, Dim_Local, Dim_Empresas e Dim_Tempo.

Dimensão	Atributos	Descrição
Local	Id_Local	Identificação da localidade
	Localidade	Nome da localidade
	Id_Posto	Identificação do Posto Administrativo
	Posto Administrativo	Nome do Posto Administrativo
	Id_Distrito	Identificação do Distrito
	Distrito	Nome do Distrito
	Id_Prov	Identificação da Província
	Província	Nome da Província

Dimensão	Atributos	Descrição
Empresa	Id_empresa	Identificação da empresa
	Nome_Emp	Nome da empresa
	Id_TipoEmpresa	Identificação do tipo da empresa
	Id_Local	Identificação da Localidade
	Ano_Constituicao	Ano de formação da empresa
	Id_CAE	Identificação do código de atividades económicas
	Id_FJR	Identificação de forma jurídica da empresa
	NUIT	Número único de identificação tributária

Dimensão	Atributos	Descrição
Tempo	Id_Tempo	Identificação do tempo
	Dia	Dia
	Mês	Mês do ano
	Trimestre	Número do trimestre
	Semestre	Número do Semestre
	Ano	Ano

Dimensão	Atributos	Descrição
Indicadores	Id_Indicador	Identificação do indicador do negócio
	Id_Industria	Identificação do indicador da indústria
	Industria	Nome do indicador da Indústria
	Id_Turismo	Identificação do indicador do turismo
	Turismo	Nome do indicador do turismo
	Id_Transporte	Identificação do indicador do Transporte
	Transporte	Nome do indicador do transporte
	Id_ComeServ	Identificação do indicador de comércio e serviços
	ComercioServico	Nome do indicador do comércio e serviços
	Id_Restaura	Identificação do indicador da restauração
	Restauracao	Nome do indicador da restauração

6.2.4. Identificação de factos

Através do levantamento de requisitos efetuado, identificou-se para este DM apenas uma tabela de facto que armazenará os dados sumarizados das estatísticas económicas e empresariais.

Facto	Atributo	Descrição
Económicas	Id_Empresa	Identificação da Empresa
	Id_Local	Identificação da localização
	Id_Indicador	Identificação do indicador do negócio
	Id_Tempo	Identificação do tempo
	Emprego	Número de emprego
	Remuneracao	Valor remunerado
	Encargos	Encargos
	Volume_Negocio	Volume de negócio

6.3. MODELAÇÃO

Depois da identificação dos processos de criação de DM, segue-se a fase de modelação do Data Mart.

6.3.1. Modelo conceptual

O modelo conceptual do DM das estatísticas económicas é constituído por quatro tabelas dimensionais e uma de facto, como se mostra nas figuras 6.3.1.1 e 6.3.1.2. Considera-se esta fase como representação de alto nível na estrutura de dados.

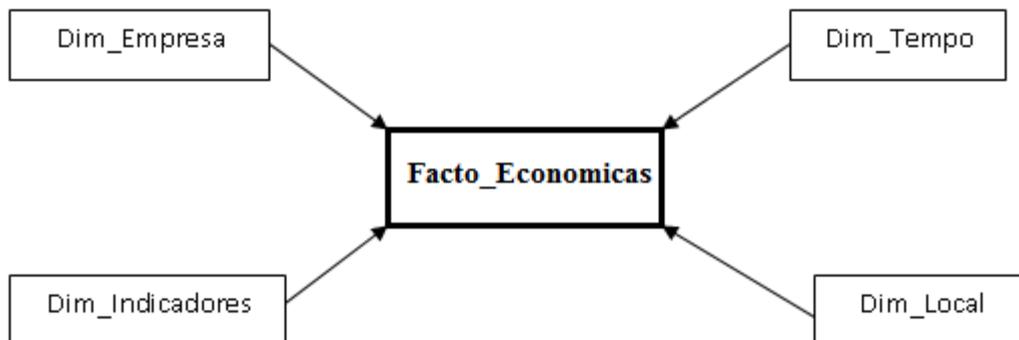


Figura 6.3.1.1 - Modelo conceitual das Estatísticas Empresariais

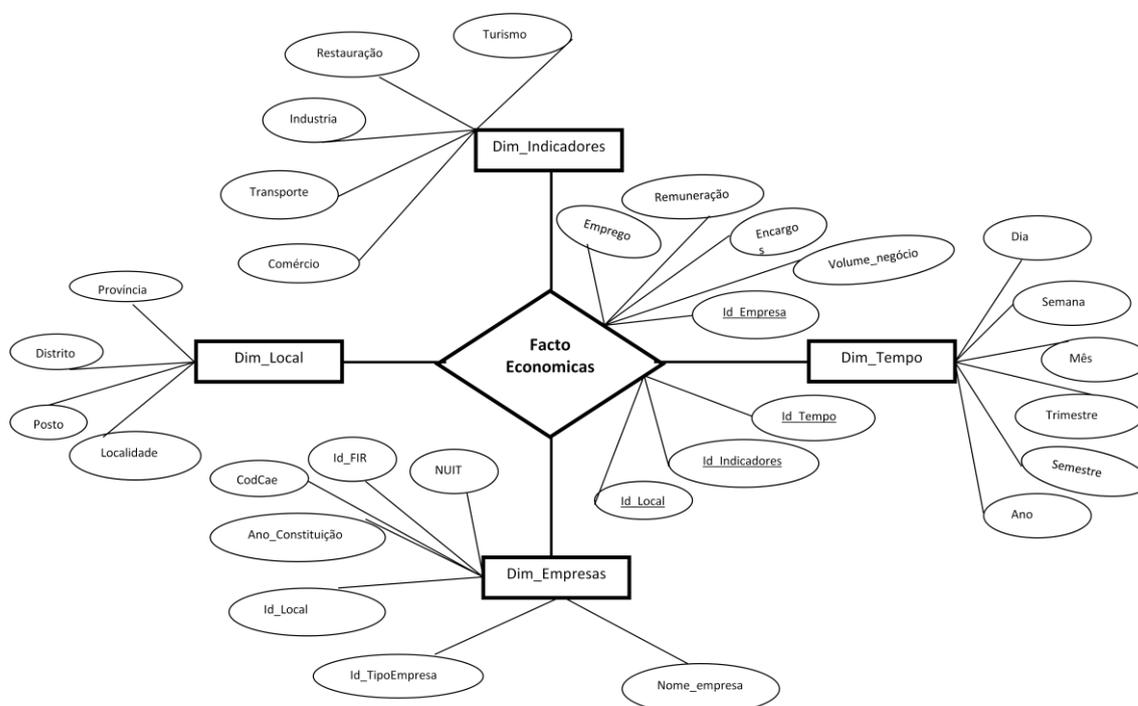


Figura 6.3.1.2 - Modelo conceitual com os atributos

6.3.2. Modelo lógico

O modelo lógico é a transformação do modelo conceitual para a forma de como a BD será concebida ou implementada em Sistemas de Gestão de Bases de Dados. As figuras abaixo ilustram essa representação, assentando na implementação em bases de dados relacionais.

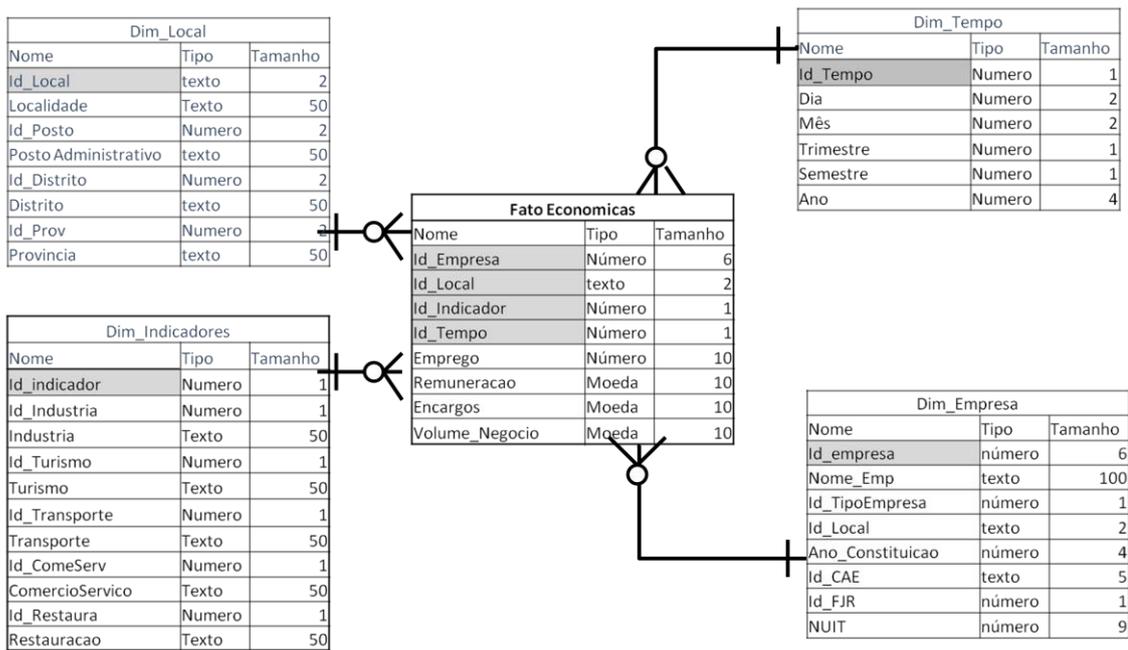


Figura 6.3.2.1 – Diagrama do Modelo lógico das estatísticas Económicas e Empresariais

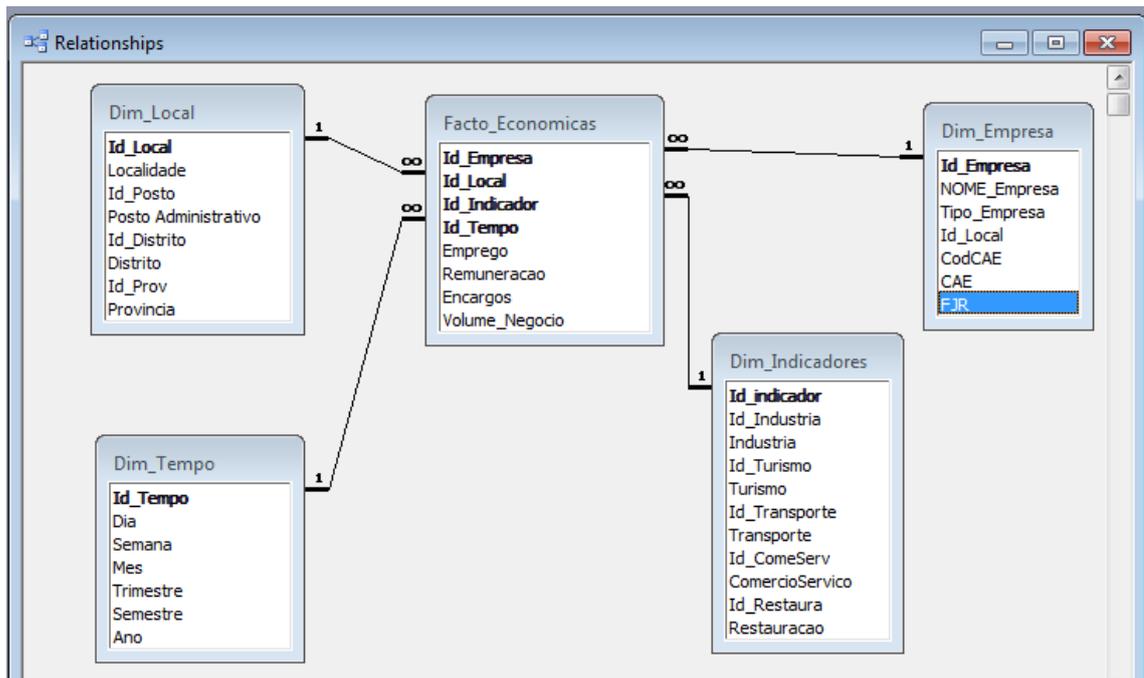


Figura 6.3.2.2 - Esquema do modelo lógico Para Estatísticas Económicas, concebido em Access



Figura 6.3.2.3 - Esquema do modelo lógico Para Estatísticas Econômicas, concebido em SQL.

6.3.3. Projeto físico

O projeto físico é a maneira exata de como é implementado o projeto lógico, faz a descrição completa da implementação da base de dados. É importante nesta fase ter a visão sobre a segurança dos dados na base.

Alguns fatores importantes que se deve considerar são:

- Particionamento
- Paralelismo no funcionamento
- Parâmetros ao configurar o Sistema de Gestão de Base de Dados (SGDB).
- Indexação

Os tópicos acima merecem um cuidado especial, devendo ser detalhados para o melhoramento da qualidade do sistema a ser implementado. É um trabalho que deve ser feito de forma contínua e por um profissional com gabarito de competência reconhecida na área das características do sistema a ser implementado.

Outros pontos importantes do DW e de aspetos não específicos totalmente na fase de elaboração, mas é através destes pontos que o projeto pode ganhar enfase de sucesso:

Metadados

Estes são conhecidos como dados que descrevem outros dados aplicáveis no ambiente do DW. Na verdade descrevem igualmente a estrutura e formato de dados. Os metadados são uteis em todos os níveis de DW, são usados para controlo e gestão do DW, servem de dicionário de dados desde o início até ao fim dos processos de conceção do modelo, processos ETL, estrutura e significado de dado, modelo de dado, isto é, toda documentação do projeto incluída.

Ao falar de documentação referimos igualmente que o metadado é um catálogo de dados contido dentro de DW, destinado a facilitar o utilizador na localização dos dados e uso dos mesmos em análises diversas.

Não é fácil encontrar um repositório de metadados que envolva todos requisitos de um DW. O ideal é usar também como alternativa os meios mínimos incorporados nas ferramentas OLAP que possibilitam a documentação dos dados, facilitando a sua localização, acesso e análise por parte dos utilizadores.

Neste estudo os metadados vão sendo criados paulatinamente ao decorrer do processo de criação do DW. Cada fase de criação e implementação que for atingida, serão detetados os metadados adequados à fase. Estes depois podem ser exportados através do *Information Catalog Manager* existente nos pacotes de *software*.

7. DISCUSSÃO DE RESULTADOS

7.1. PROJETO DO SISTEMA DE DW NO INE-M

A construção de DW para INE teve como primeira etapa a identificação dos utilizadores e levantamento de requisitos. Com a identificação dos potenciais utilizadores, um irá gerir o projeto aquando sua implantação. Este gestor será apenas responsável pelo projeto e não pelo DW como sistema. A gerência do projeto tem um prazo de existência e é tarefa do responsável estabelecer o plano geral só do projeto e que deve ser de conhecimento de todos os elementos que compõe a equipa de desenvolvimento.

No plano geral do projeto, deve conter os recursos disponíveis para o projeto, prazo do projeto e a expectativa dos utilizadores perante o sistema a ser iniciado. A identificação do responsável pela gestão do projeto tem sido o maior problema para este tipo de atividade.

O gestor de DW enquanto sistema é permanente. Enquanto durar o ciclo de vida do sistema, deve preocupar-se com a execução dos processos do DW.

Para este estudo, pretende-se um sistema automatizado que possa gerar dados analíticos, eliminando os problemas que a instituição enfrenta com as aplicações operacionais ainda em uso no INE-M.

Para conceção deste projeto, contou com as ferramentas da Microsoft existentes na instituição, em particular o SQL Server 2008 instalado num dos servidores.

7.2. PARTICIPAÇÃO, LOCALIZAÇÃO E CARGA

A elaboração deste trabalho contou com a participação dos gestores estratégicos do INE – M. Estes facilitaram o trabalho de levantamento de requisitos na instituição, e também demonstraram interesse e entusiasmo em verem esta ferramenta de apoio à tomada de decisões instalada e a funcionar na instituição. Atendendo a este interesse dos gestores, pode concluir-se que, o risco de fracasso do projeto no INE-M será muito menor.

O projeto estará ao cargo do departamento de informática da instituição, local onde se encontram os servidores onde o DW será instalado. Os utilizadores finais terão acesso de princípio via Web, através dos códigos de restrição que serão atribuídos em conformidade com o nível estratégico de gestão. Existe uma expectativa futura de migração para o BI e *dataminig*.

Na primeira fase da carga dos DMs estará a cargo do técnico de cada setor do pelouro, que até aqui é responsável pelos dados. Este técnico terá uma assistência direta de um técnico de informática, que será indicado para essa finalidade, e que tenha feito parte da equipa de desenvolvimento.

7.3. ATUALIZAÇÃO E MANUTENÇÃO DO DW

A ferramenta de desenvolvimento de DW deve estar preparada para suportar qualquer atualização ou manutenção. Para o efeito deve contar com atividades tais como:

1. Definição da periodicidade da atualização.
2. Possibilidade de interagir com outras ferramentas.
3. Automatização do processo ETL.

A manutenção estará a cargo da equipa responsável pela carga do DM e com supervisão contínua de um diretor de cada pelouro.

7.4. CONSIDERAÇÃO FINAL

Com este trabalho, espera-se ter ajudado ao estudo de caso, com uma ferramenta apropriada de apoio à tomada de decisão ágil e segura; Ter propiciado uma ferramenta poderosa, na redução de tempo de resposta tanto ao nível interno como externo; Ter facilitado a criação e consulta de relatórios com apoio de ferramentas *front-end*, assim como ter contribuído para a melhoria de qualidade de dados e da informação.

8. CONCLUSÕES

Com o objetivo de alcançar os objetivos do estudo, os resultados aqui relatados são fidedignos, seguindo a convicção de assim permitir um estudo mais aprofundado do tema. Sendo este um trabalho de investigação, acredita-se que os resultados irão ajudar o INE-M, na resolução dos seus problemas. Este tema oferece muita oportunidade de investigação como forma de elevar o nível académico.

O estudo permitiu identificar a melhor metodologia e modelo de DW para apoio estratégico ao INE-M, como entidade única de produção das estatísticas oficiais em Moçambique.

Foi através de várias opiniões colhidas em muitos autores por meio de levantamento bibliográfico e pelos utilizadores finais, que foi possível escolher das metodologias e modelos, a melhor que tivesse boa aplicabilidade para a realidade moçambicana, concretamente numa instituição pública de produção e disseminação de estatísticas oficiais no País.

Como se afirmou acima, o INE-M não possui nenhum modelo analítico de dados, quando este for bem implementado segundo as propostas, será um ganho considerável para a instituição e para o País.

Considera-se particularmente positivo que os pontos aqui propostos que podem efetivamente orientar a elaboração de um DW no INE – M, onde se destacam as fases de construção do projeto de DW. Donde a partir deste projeto, pretende-se um repositório de dados estatísticos integrados, coerentes, estruturados, e de fácil acesso, que facilite análises complexas e responda a todos os interesses da instituição.

Através desta ferramenta os gestores estratégicos podem administrar o seu pessoal por intermédio de competências, permitindo uma boa formação de equipas de partilha de conhecimentos.

Por meio dos indicadores, permite um acompanhamento eficaz do negócio.

Com este trabalho de dissertação adquiriu-se experiência de pesquisa e na conceção de DW em particular. Estamos convictos que o mesmo ajudará muitos pesquisadores e profissionais da área dado que encontra-se bem documentado.

9. LIMITAÇÕES E RECOMENDAÇÕES PARA TRABALHOS FUTUROS

9.1. LIMITAÇÕES NO ESTUDO

Atualmente falar de DW não constitui novidade. O difícil é encontrar nas empresas, técnicos qualificados para construir ou implementar uma ferramenta desta natureza.

Outra limitação é a identificação da melhor metodologia que na prática pode servir de apoio estratégico na tomada de decisões na organização.

As atividades do dia-a-dia das instituições estão em constante movimentação, o que dificulta conseguir tempo livre ou suficiente dos responsáveis das instituições, que possam ajudar durante a pesquisa para transmitir os seus conhecimentos, e fornecer ideias ou pareceres sobre a instituição que dirigem.

Apesar da abertura total dos responsáveis do INE-M, foi notório que existiu alguma cautela na disponibilização dos dados para este trabalho, ou seja, aquilo a que se chama sigilo ou segredo estatístico.

9.2. RECOMENDAÇÕES PARA TRABALHOS FUTUROS

O trabalho desta natureza é importante para todos profissionais de área de desenvolvimento de sistemas. Recomenda-se a criação de um “dicionário” para além de metadados existentes, para melhor auxílio no uso do DW e de outros sistemas.

A metodologia escolhida de DMs incrementais é recomendável para todas organizações ou instituições públicas semelhantes ao INE-M, que financeiramente depende do orçamento geral do estado ou de doações dos parceiros de negócios.

De seguida deve procurar-se dotar o sistema com ferramentas de código aberto, que seja de uso livre mas com grandes potencialidades no desempenho.

Quando implementado o primeiro DM do sistema, há que procurar junto dos utentes, conhecer a sua satisfação, avaliando o tempo de respostas do sistema, fazendo as respetivas correções ou retificações nos módulos que demonstrarem deficiências, para que os próximos erros sejam mínimos ou inexistentes.

Sendo o INE-M uma instituição que trabalha em cooperação com outras instituições pertencentes ao grupo de Sistema Estatístico Nacional deve padronizar o modelo para facilitar o seu desempenho.

O departamento de informática e sistemas de informação (DISI), deverá procurar conhecer e implementar o uso de técnicas mais recentes, como por exemplo, o DMg para encontrar outras potencialidades do sistema DW.

10. BIBLIOGRAFIA

- Abreu, F. S. G. G (2008), Desmistificando o conceito de ETL, Revista de Sistemas de Informação. URL: http://www.fsma.edu.br/si/artigos/v2_artigo1.pdf
- Araújo, E. M. T. (SD), URL: <http://www.devmedia.com.br/um-estudo-sobre-as-ferramentas-olap/6691>
- Caldeira, C. P. (2008). Data Warehousing – Conceitos e modelos
- Carlos, F. M. M. (2008). Data Warehouse, Num Instituto Politécnico
URL:http://repositorio.utad.pt/bitstream/10348/426/1/msc_fmcarlos.pdf
- Carvalho, I. M.; Muñhoz, D. L.C.; Ferreira, M. V. A. S.; Santos, P. M.; Rover, A. J.; Fialho, F. A. P. (nd). Contribuições das tecnologias como ferramentas de gestão de conhecimento.
URL:http://www.egov.ufsc.br/portal/sites/default/files/artigo_1_conegov.pdf
- Crisp (2007). URL: [http:// www.markosweb.com/www/crisp-dm.org/Em cache; visitado em Setembro 2012.](http://www.markosweb.com/www/crisp-dm.org/Em_cache;visitado_em_Setembro_2012)
- Dill, S. L. (2002). Metodologia para o Desenvolvimento de Data Warehouse. URL: <http://homes.dcc.ufba.br/~mauricio052/Material%20Artigo/Uma%20Metodologia%20para%20Desenvolvimento%20de%20Data%20Warehouse%20e%20Estudo%20de%20Caso.pdf>
- Facundes, E. M.(); URL: www.efagundes.com
- Fayyad, U. M.; PIATETSKY-SHAPIRO, G.; SMYTH, P. (1996). From Data Mining to Knowledge Discovery: An Overview. Knowledge Discovery and Data Mining, Menlo Park : AAAI Press.
- Gouveia, L. M. B. (1996). Sistemas de Informação. URL: http://homepage.ufp.pt/lmbg/textos/si_texto.pdf.
- Gouveia, L. M. B. (2004). Sistemas de Informação de Apoio à Gestão; Spi.
- Hossain, L.; Patrick, J. D. & Rashid, M. A. (2002). Enterprise Resource Planning: Global Opportunities and Challenges. Sydney: IGP.
- INE - Moçambique (2009). Organograma do INE.
URL:http://www.ine.gov.mz/home_page/o_ine/dir_organograma/organograma
- Inmon, W. H. (1997). *Como construir o data warehouse*. (2ª ed.) Rio de Janeiro: Editora Campus
- Inmon, W. H. (2002). Building the Data Warehouse. Third Edition. Wiley
- Julião, R.P. (2009). Apontamentos da disciplina de SIG nas Organizações. Mestrado em Ciência e Sistemas de Informação Geográfica no Instituto Superior de Estatística e Gestão de Informação da Universidade Nova de Lisboa, Lisboa.

- Kimball, R.; Ross, M. (2002). *The Data Warehouse Toolkit. Second Edition: The Complete Guide to Dimensional Modeling*. Wiley
- Knuth, D. E. (sd). *Selected Papers on Computer Science*. Cambridge: Cambridge University Press. 274 p. p. 1-2. ISBN 1-1881526-91-7
URL:http://pt.wikipedia.org/wiki/Informa%C3%A7%C3%A3o#cite_note-knuth-12
- Laudon, K. C., & Laudon, J. P. (2003). *Management Information Systems (8 ed.)*. New Jersey: Prentice Hall.
- Lustosa, R. B. (2009). *Processo de Desenvolvimento Participativo de Sistemas de DW*.
URL:http://www.ccsa.ufpb.br/ppga/site/arquivos/dissertacoes/dissertacao_61.pdf
- Machado, F. N. R. (2000). *Projeto de Data Warehouse. Uma visão multidimensional*. São Paulo: Editora Érica,
- Machado, F. N. R. (2006). *Tecnologia e projeto de Data Warehouse*. São Paulo: Editora Érica
- Perkins, A. (2000). *Critical Success Factors for Data Warehouse Engineering*;
<http://www.tdan.com/view-articles/4836>.
- Quaresma, S.; Perpetuo, F. (2004). *Modelação Conceptual de Data Warehouses para (Instituto Nacional de Estatística, Portugal) "data warehousing in statistical production"*. Lisboa.
- Scalabrin, E. E. (SD). *Abordagem de desenvolvimento de data warehouse*.
URL:<http://www.ppgia.pucpr.br/~scalabrin> (consultada em 05/03/2012).
- Santos, A., (2008). *Gestão Estratégica - Conceitos, modelos e instrumentos*. Lisboa. Escolar Editora. SAP Portugal, 2010. SAP Business Intelligence.
URL:<http://www.sap.com/portugal/platform/netweaver/businessintelligence/index.epx>
- Santos, I. M. (2003). *Data Warehouse como ferramenta de auxílio em sistemas de monitoramento ambiental*. Cuiabá-MT Brasil. URL:
<http://www.ufmt.br/cacomp/Downloads/monografias/DW-AcompanhamentoAmbiental.pdf>
- Silva, C. I. M. e (2011). *Desenvolvimento de Software e Sistemas Interativos*.
URL:<http://repositorio.ipcb.pt/browse?type=author&value=Silva%2C+Cristiana+I+sabel+Morgado+e>
- Silva, M. M. da, (2003). *Integração de sistemas de Informação*, FCA – Editora informática, ISBN 972-722-391-5.
- Singh, H. S. (2001). *Data warehouse – conceitos, tecnologias, implementação e gerenciamento*. São Paulo: Makron Books.

- Takai, O.K.; Italiano, I.C.; Ferreira, J.E. (2005). Introdução a Banco de Dados. [URL:http://www.ime.usp.br/~jef/apostila.pdf](http://www.ime.usp.br/~jef/apostila.pdf)
- Teixeira, P. A. M.S. (2005). Estudo da adoção do Local e-government. [URL:http://homepage.ufp.pt/lmbg/monografias/mscpteixeira05.pdf](http://homepage.ufp.pt/lmbg/monografias/mscpteixeira05.pdf)
- Varajão, J. (2005). Arquitectura da Gestão de Sistemas da Informação, editor FCA - Brasil, ISBN: 9727225071, 9789727225071.
- Ziulkoski, L.C.C., (2003). Técnicas de Aquisição de conhecimento. [URL:http://www.inf.ufrgs.br/bdi/wp-content/uploads/Monografia-LuisZiulkoski.pdf](http://www.inf.ufrgs.br/bdi/wp-content/uploads/Monografia-LuisZiulkoski.pdf)

11.ANEXOS

 <p>INSTITUTO NACIONAL DE ESTATÍSTICA Direcção de Estatísticas Sectoriais e de Empresas</p> <p>Inquérito Mensal ao Comércio e Serviços Mês de referência: _____ de 200__</p> <p><small>COLABORE! FORNEÇA A INFORMAÇÃO SOLICITADA DENTRO DO PRAZO. VEJA BEM AS INSTRUÇÕES ANTES DE COMEÇAR A PREENCHER.</small></p>		<p>Instrumento de notação do Sistema Estatístico Nacional Lei 7/96 Julho de 1996, de <u>resposta obrigatória</u> Registrado no INE sob o Nº IMCS - 04</p> <p>Devolva o questionário devidamente preenchido no prazo de <u>5 dias após o mês de referência</u> à Delegação Provincial do INE.</p> <p>A confidencialidade dos dados individuais é garantida por Lei</p>					
A. Identificação e localização do estabelecimento	Nome:						
	Localização: Av., Rua ou Praceta:						
	Província:		Distrito:				
	Telefone:	Fax:	Email:				
B. Actividade	Sector de Actividade:		COMÉRCIO <input type="checkbox"/>		PRESTAÇÃO DE SERVIÇOS <input type="checkbox"/>		
	Actividade principal:						
	Outras actividades:						
C. Vendas e/ou Prestação de Serviços		Valor Total de Receitas Sem Impostos (MT) →					
D. Mão de Obra ao Serviço durante o mês			Homens		Mulheres		Total
			Nacionais	Estrang.	Nacionais	Estrang.	
	Número de Pessoas ao Serviço						
	Remunerações mensais (MT)						
	Número de horas trabalhadas						
	Encargos Sociais obrigatórios (MT)						
	Encargos Sociais Facultativos (MT)						
Encargos na Formação (MT)							
E. Observações	Incluir qualquer esclarecimento relativo ao mês de referência que julgar pertinente, em particular sobre a mudança de sector de actividade ou de actividade principal.						

Anexo 11.1 - Ficha de dados área de Comércio e Serviços; Fonte: INE – M.

**Guião de apoio à entrevistas
Gestão de topo**

Estimado(a) senhor(a), serve este meio, para pedir algum tempo da sua disponibilidade para esta pequena entrevista. Os dados aqui recolhidos serão guardados sigilosamente e destinam-se única e exclusivamente para o estudo de um modelo conceptual de Data Warehouse a propor para INE – M, como tema de dissertação. Desde já, agradeço a colaboração.

I.1. Dados Gerais

I.1. Ano:2012; Mês: _____; Departamento / Direção / Presidência: _____;

I.2. Dados Pessoais

I.2.1 - Nome (Opcional):

I.2.2 - Género (marque com "X"): 1. Homem _____; 2. Mulher _____

II. Questões gerais (sublinhar, marcar X ou descrever)

II.1- Há quantos anos trabalha na instituição: A: 0 – 5; B: 6 – 10; C: 11 – 15; D: 16-20; E: 21 ou +

H.2- Qual é o nível de escolaridade que terminou: 1 – Primário; 2-Secundário; 3-Médio; 4-superior

H.3- Há quanto tempo (anos) é gestor: A: 0 – 5; B: 6 – 10; C: 11 – 15; D: 16-20; E: 21 ou +

II.4 – Na qualidade de gestor usa algum software específico de gestão? A: Sim, B:Não

II.5- Qual é o principal objetivo da sua instituição? _____

II.6- Como mede o cumprimento dos objetivos? _____

II.7- Quem são os principais clientes da empresa? _____

II.8- Como gestor considera o uso das novas tecnologias da informação importante na sua atividade? A:Sim; B:Não; C:Sem opinião

O objetivo principal do DW é apoiar aos gestores na tomada de decisão, aceita ou apoia a implementação deste instrumento nesta instituição? A: Sim; B:Não; C:Sem opinião

Observação / sugestão:

Obrigado pela atenção dispensada e votos de bom trabalho.

Anexo 11.3 - Guião de apoio na recolha de requisitos (gestão de topo)

**Guião de apoio à entrevista
Gestão Intermédia**

Estimado(a) senhor(a), serve este meio, para pedir algum tempo da sua disponibilidade para esta pequena entrevista. Os dados aqui recolhidos serão guardados sigilosamente e destinam-se única e exclusivamente para o estudo de um modelo conceptual de Data Warehouse a propor para INE – M, como tema de dissertação. Desde já, agradeço a colaboração.

I.1. Dados Gerais

I.1. Ano:2012; Mês: _____; Departamento / Direção / Presidência: _____;

I.2. Dados Pessoais

I.2.1 - Nome (Opcional): _____

I.2.2 - Género (marque com “X”): 1. Homem _____; 2. Mulher _____

II. Questões gerais (sublinhar, marcar X ou descrever)

II.1- Há quantos anos trabalha na empresa: A: 0 – 5; B: 6 – 10; C: 11 – 15; D: 16-20; E: 21 ou +

H.2- Qual é o nível de escolaridade que terminou: 1 – Primário; 2-Secundário; 3-Médio; 4-superior

H.3- Há quanto tempo (anos) é gestor: A: 0 – 5; B: 6 – 10; C: 11 – 15; D: 16-20; E: 21 ou +

II.4 – Na qualidade de gestor usa algum software específico de gestão? A: Sim, B:Não

II.5- Qual é o principal objetivo da sua Direção? _____

II.6- Como mede o cumprimento dos objetivos? _____

II.7- Quem fornece dados para sua Direção? A: Setor Formal; B: Setor Informal; C: Outros

II.8- Como é feita a recolha de dados no seu setor? A- Entrevista; B- Questionário; C- Outros meios

II.9- Qual é a periodicidade de recolha de dados no seu setor? A: Diário; B:Semanal; C:Mensal; D: Trimestral; E: Semestral; F: Anual.

II.10 – Os dados recolhidos já vêm uniformizados? A: Sim; B:Não; C: Não sabe

II.11 – Com estes dados consegue responder pontualmente os interesses da instituição? A: Sim; B: Não; C: Não sempre

II.12- O objetivo principal do DW é apoiar os gestores na tomada de decisão, aceita ou apoia a implementação deste instrumento nesta instituição? A: Sim; B:Não; C:Sem opinião

Observação / sugestão:

Obrigado pela atenção dispensada e votos de bom trabalho.

Anexo 11.4 - Guião de apoio na recolha de requisitos (gestão intermédia)

Guião de apoio a entrevistas Informática

Estimado(a) senhor(a), serve este meio, para pedir algum tempo da sua disponibilidade para esta pequena entrevista. Os dados aqui recolhidos serão guardados sigilosamente e destinam-se única e exclusivamente para o estudo de um modelo conceptual de Data Warehouse a propor para INE – M, como tema de dissertação. Desde já, agradeço a colaboração.

I.1. Dados Gerais

I.1. Ano:2012; Mês: _____; Departamento / Direção /Presidência: _____;

I.2. Dados Pessoais

I.2.1 - Nome (Opcional):

I.2.2 - Género (marque com "X"): 1. Homem _____; 2. Mulher _____

II. Questões gerais (sublinhar, marcar X ou descrever)

II.1- Há quantos anos trabalha na instituição: A: 0 – 5; B: 6 – 10; C: 11 – 15; D: 16-20; E: 21 ou +

H.2- Qual é o nível de escolaridade que terminou: 1 – Primário; 2-Secundário; 3-Médio; 4-superior

H.3- Há quanto tempo (anos) é chefe do Departamento: A: 0 – 5; B: 6 – 10; C: 11 – 15; D: 16-20; E: 21 ou +

II.4- Qual é a principal tarefa do Departamento? _____

II.5- Quantas aplicações a instituição possui e onde funcionam?

II.6- Quantos servidores existem na instituição?

II.7- Que tipo de softwares e aplicações a instituição usa?

II.8- Das aplicações existentes possuem documentação técnica?

II.9- O seu departamento possui condições técnicas de conceção e desenvolvimento de base de dados? A- Sim; B- Não

II.10- O objetivo principal do DW é apoiar aos gestores na tomada de decisão, O departamento está capacitado para o desafio? A: Sim; B:Não; C:Sem opinião

Observação / sugestão:

Obrigado pela atenção dispensada e votos de bom trabalho.

Anexo 11.5 - Guião de apoio na recolha de requisitos (Departamento de Informática)