

DANIEL EDUARDO DURÃES AGUIAR

***CONFIABILIDADE DA INFORMAÇÃO:***

A IMPORTÂNCIA DESTE ATRIBUTO NA QUALIDADE DA INFORMAÇÃO  
NAS ORGANIZAÇÕES

BELO HORIZONTE

2014

DANIEL EDUARDO DURÃES AGUIAR

***CONFIABILIDADE DA INFORMAÇÃO:***

A IMPORTÂNCIA DESTE ATRIBUTO NA QUALIDADE DA INFORMAÇÃO  
NAS ORGANIZAÇÕES

Monografia apresentada ao fim da especialização em Gestão Estratégica da Informação na Universidade Federal de Minas Gerais como requisito parcial para obtenção do título de Especialista em Gestão Estratégica da Informação.

Orientador: Prof. Eduardo Ribeiro Felipe

BELO HORIZONTE

2014

## DEDICATÓRIA

Dedico esta monografia a minha mãe e ao meu pai (*in memoriam*) que sempre acreditaram e incessantemente me incentivaram juntamente ao meu irmão, a minha namorada por toda a ajuda, o apoio e carinho.

## AGRADECIMENTOS

Agradeço a Deus por mais esta meta alcançada e por me proporcionar este momento, ao meu orientador Eduardo Ribeiro Felipe pela orientação e auxílio, e a todos que de alguma maneira desejam o meu sucesso.

## RESUMO

O trabalho realiza uma investigação sobre a confiabilidade de informações a nível gerencial em uma organização do ramo de telecomunicações, trabalhando a condição de credibilidade das informações em diversos níveis de acesso. Dentro deste cenário foi proposto o desenvolvimento e implementação de uma ferramenta de tratamento de dados e divulgação dos indicadores de desempenho com o acesso por meio de um navegador de internet, sendo uma solução para eliminar a desconfiança sobre as fontes provedoras de informação. Este trabalho promove uma análise dos conceitos usados pela arquitetura BI e ainda abordam seus atributos, conceitos de informação, métricas para medição da confiabilidade informacional, além de mostrar a possibilidade de uma solução escalar, utilizando ferramentas usadas para tais fins, desde que o escopo e uso sejam bem definidos.

**Palavras chave:** *Confiabilidade; Informação; Atributos da Informação; Qualidade da Informação; Business Intelligence; Data Warehouse; Microsoft Excel; MySQL; PHP.*

## LISTA DE ILUSTRAÇÕES

Figura 1 - Os níveis hierárquicos da informação	15
Figura 2 - Componentes do ambiente de <i>Business Intelligence</i>	19
Figura 3 – Diagrama de uma ferramenta de ETL convencional	21
Figura 4 – Elementos Básicos de um <i>Data Warehouse</i>	23
Figura 5 – <i>Data Marts</i> em um <i>Data Warehouse</i>	25
Figura 6 – Representação de um cubo OLAP	29
Figura 7 – Representação de um modelo <i>Star Schema</i>	30
Figura 8 – Representação de um modelo Snowflake	31
Figura 9 – Uso do esquema <i>Snowflake</i> na diagramação das tabelas	39
Figura 10 – Fluxograma da ferramenta de ETL	41
Figura 11 – Fluxo de processos da ferramenta ETL desenvolvida	43

## LISTA DE QUADRO

Quadro 1 – Dimensões e atributos da informação	16
Quadro 2– Dimensão da informação por processo de BI	43
Quadro 3 – Comparativo das dimensões da informação analisadas	45

## LISTA DE ABREVIATURAS E SIGLAS

BD – Banco de Dados.

DW – *Data Warehouse* – ou simplesmente Armazém de dados.

DM – *Data Mart* – Subconjunto de um *Data Warehouse*.

ETL – *Extract, Transform e Load* – Processo de extração, transformação e carga dos dados.

OLAP – *On-Line Analytical Processing*.

PHP - *PHP is Hypertext Processor* (PHP é um processador de hipertexto).

SGBD – Sistema Gerenciador de Banco de Dados.

SQL – *Structured Query Language* – linguagem estruturada de dados.

VBA – *Visual Basic for Applications*.

## SUMÁRIO

1.	INTRODUÇÃO	09
1.1.	Problema	11
1.2.	Objetivo Geral	11
1.3.	Justificativa	12
1.4.	Estrutura do Trabalho	12
2.	REFERENCIAL TEORICO	13
2.1	Fluxo cíclico de dado, informação e conhecimento	13
2.2	Qualidade da informação	16
2.2.1.	Medição da qualidade da informação	16
2.2.2.	Dimensão e atributos relacionados à informação	17
2.3	Modelagem Multidimensional	18
2.3.1.	ETL	19
2.3.2.	Data Warehouse	21
2.3.3.	Data Mart	24
2.3.4.	OLAP	25
2.4.	Modelagem Multidimensional	28
2.5.	Estrutura do Projeto	31
3.	METODOLOGIA	32
3.1	Desenvolvimento	34
3.1.1	Iniciação	35
3.1.2	Planejamento	35
3.1.3	Execução	36
3.1.4	Monitoramento e Controle	36
3.1.5	Encerramento	37
3.2	Modelagem dos dados	38
3.3	Processo de Extração, Transformação e Carga (ETL)	40

3.4	OLAP	41
4.	CONCLUSÃO	44
	GLOSSÁRIO	51

## 1. INTRODUÇÃO

A velocidade em que ocorrem transformações em âmbito global, principalmente no que diz respeito aos negócios das organizações envolvidas, se torna cada vez mais veloz, assim as empresas buscam e necessitam cada vez mais informações geradas a partir de dados coletados em seus processos de negócio (MORESI, 2000).

Existe o consenso de que na sociedade pós-industrial, cuja economia assume tendências globais, a informação passou a ser considerado um precioso capital (MORESI, 2000). Há também o consenso de que estamos em uma geração em que a informação se tornou o principal criador de riquezas (TOFLER, 2001). De modo que a informação, algo abstrato, passou a se equiparar aos recursos de produção, materiais e financeiros. Compilando as informações analisadas, junto ao conhecimento sobre o negócio do gestor este volume informacional propicia uma tomada de decisão mais assertiva para manter a empresa mais competitiva no mercado (MORESI, 2000).

Uma empresa pode gerar uma quantidade de dados que seria inimaginável á tempos atrás em um único dia, porém muitas delas não são coletadas, ou são ignoradas. Sendo que estes dados às vezes podem ser de grande valia quando formatados e tratados de forma certa, eles podem proporcionar uma visão antes não percebida e influência beneficemente a decisão de um gestor. Além de ser extremamente sensível a percepção de que os maiores ativos das organizações estão se tornando e já são em muitas delas bens intangíveis. SANTOS (2006) descreve em seu artigo estes ativos não visíveis são parte fundamental para a manutenção competitiva da organização.

SORDI (2008) ressalta a importância da necessidade de uma informação de qualidade, onde não se adianta ter ferramentas tecnologicamente avançadas, porém, sem uma fonte de dados clara e confiável para extrair uma informação de qualidade.

Os indicadores são originários na área econômica e eram basicamente exclusivos para fins quantitativos, a partir dos anos 1920 começou a se difundir os indicadores com foco social, quando houve a percepção da limitação dos indicadores econômicos (SOLIGO, 2012). Os administradores passaram a reconhecer a necessidade e a utilidade de indicadores que não fossem financeiros, que estes podiam ser utilizados de forma estratégica já que estes permitem a visão de uma tendência, útil para um planejamento á longo prazo (ECCLES e MAVRINAC, 1995). Contudo Dempsey et al. (1997), afirma que as questões de acesso e uso aos indicadores não financeiros são limitados e que a busca por informação deverá ir até o ponto em que os seus custos de aquisição se igualar aos seus benefícios, ou seja, que o produto de informação destes indicadores justifique a sua formulação. Os indicadores de desempenho e qualidade assim como qualquer outra informação dentro da organização devem originar de uma fonte segura para aqueles que devem consumir a informação. Os resultados e números, publicados aos consumidores necessitam de estar em uma formatação visual de modo que possam demonstrar ao gestor a atual situação da produtividade ou qualidade, de seus produtos ou serviços. Seguindo as ideias sobre indicadores de ATKINSON (1998), os indicadores devem auxiliar na gestão, focando em pontos críticos para o sucesso no negócio. Assim com a influência das ferramentas de suporte a decisão, o gestor possui uma maior probabilidade de acerto em suas decisões, visto que os indicadores procuram dar uma visão mais clara do negócio.

No trabalho serão expostos os requisitos para se preencher satisfatoriamente os atributos de uma informação de qualidade segundo SORDI (2008) dando ênfase aos que são ligados a confiabilidade e como foram atendidos. Será apresentada também a teoria de *Business Intelligence*, além de mostrar o *Microsoft Excel*, em um cenário real, usado como ferramenta de coleta e limpeza de dados e sua interação com um *Data Warehouse*, este feito usando o sistema gerenciador de banco de dados gratuito *MySQL*.

## 1.1. Problema

A confiabilidade das informações corporativas é uma premissa para tomadas de decisões mais seguras. A fragilidade deste cenário expõe a empresa a riscos, podendo induzir ao erro, causando conflitos e dúvidas entre os gestores, o que pode aumentar o tempo para uma tomada de decisão ou até levar a não tomar a melhor decisão quando isso se faz necessário a todo instante nas organizações e que é vital para o sucesso da empresa.

A credibilidade informacional em uma organização, em um cenário em que as mesmas não estavam suficientemente confiáveis para consumo dos gestores tomadores de decisão, podendo induzir ao erro e causando conflitos e dúvidas entre os *Stakeholders*, devido á baixa confiança e por ter aspectos contraditórios causados pela inconsistência dos dados e informações levantadas e usadas como base para a tomada daquela decisão.

Diante desta situação, seria necessário aplicar uma solução que elimine qualquer questionamento em relação à confiabilidade dos dados e informações consumidas. A criação de um sistema de informação, para centralizar informações, poderia estabelecer a confiabilidade dos dados e informações consumidas?

## 1.2. Objetivo Geral

Demonstrar, através de um estudo de caso, a importância da confiabilidade dos dados organizacionais de modo que atinja um nível de qualidade satisfatória perante aos consumidores da informação para a tomada de decisão.

Criar um sistema de informação que garanta a confiabilidade das informações e seja sem custos, utilizando técnicas de BI e Excel como coletador de dados.

Expor os conceitos de qualidade informacional de acordo com SORDI (2008) e como os atributos relacionados à confiabilidade fossem atendidos.

### 1.3. Justificativa

Hoje é fundamental para as empresas possuírem as informações mais coesas e confiáveis no menor tempo possível. Esta necessidade ocorre com cada vez mais frequência e intensidade, visto a velocidade com que o ambiente corporativo opera, principalmente com o advento da internet nas empresas e toda infraestrutura de tecnologia que trás consigo. Com o advento da rede mundial de computadores nas empresas, ocorre o aceleração de buscas e solicitações a resultados quase instantâneos, tendo deste modo inúmeras aplicações e possibilidades de automação, acelerando e melhorando os processos das empresas. Além disto, os sistemas de informação são um dos maiores fatores que influenciam uma tomada de decisão dos gestores junto com o sentimento em relação ao momento e conhecimento tácito dos mesmos. Com isto foi identificada a necessidade de se implantar um sistema de informação, para reduzir o tempo de processamento para publicar a informação e que esta seja confiável.

### 1.4. Estrutura do Trabalho

A seção 2 tem como proposta, explicar as definições usadas no trabalho para dado, informação e conhecimento. Os atributos relacionados á qualidade da informação e *Business Intelligence*.

Na seção 3 é mostrado o emprego dos conceitos relacionados á qualidade da informação para a produção confiável dos relatórios as técnicas de Business Intelligence usadas para prover o mesmo, as ferramentas usadas no projeto, suas características e sua correlação nos processos de *Business Intelligence* ao qual foi empregada e também são apresentadas as etapas que foram seguidas para a elaboração do projeto mostrado neste estudo de caso. São expostos em detalhes, o uso das técnicas de qualidade da informação e as ferramentas usadas para realizar os processos.

E na seção 4 são informados os resultados alcançados e as considerações finais.

## 2. REFERENCIAL TEORICO

### 2.1 Fluxo cíclico de dado, informação, conhecimento e inteligência.

Com a quantidade de informações geradas, e captadas pelas empresas, se torna importante o tratamento dos dados, com o intuito de gerar a informação desejada que atenda ao usuário solicitante de modo que possa influenciar positivamente a tomada de decisão. Entretanto, antes de formular como serão coletado os dados e expostas às informações perante o usuário, é preciso definir o que é entendido como dado e o método no qual é concebida a informação, pois ao elucidar ambos, o resultado da transformação de dado em informação seja o mais coeso possível. No contexto do trabalho, é definido como dado uma seqüência de símbolos quantificados ou quantificáveis (SETZER, 2001). Então pode ser usado como exemplo prático de dado, um número ou um texto.

Já informação é uma abstração que está na mente de alguém, representando algo significativo para essa pessoa, sendo que esta interpretação pode mudar de pessoa para pessoa, e que não pode ser armazenado em um computador, somente o dado (SETZER, 2001). Além disso, os dados gravados em um sistema computacional, de acordo com a forma em que são mostradas, podem variar a informação assimilada de pessoa para pessoa que veja o resultado exibido, baseando no conhecimento prévio que ela possua (SETZER, 2001).

A informação em qualquer ponto de vista seja ele computacional ou não, também pode ser entendida como uma representação de fatos ou ideias que podem alterar o estado de conhecimento do individuo (CARVALHO *apud* GONÇALVES, 1996). E que ao interpretar não deve ser analisado somente a informação em si como um pedaço ou fragmento, mas também o contexto de onde ela foi extraída (CAPRA, 1996).

Segundo MORESI (2000), No ambiente corporativo, na sociedade pós-industrial, a informação não é um recurso, mas “o” recurso. E a aceitação desta ideia coloca como o recurso chave de competitividade efetiva, de diferencial de mercado e de lucratividade na sociedade moderna.

SETZER (2001) define conhecimento, como uma abstração, pessoal, de algo que foi experimentado, vivenciado. Pode ser associada a uma informação, junto à experiência de vida de uma pessoa, porém enquanto a informação pode ser conceitual ou prática, o conhecimento sempre tende ser prático.

O conhecimento pode ser alterado dependendo das informações recebidas (CARVALHO *apud* GONÇALVES, 1996), já que o indivíduo interage com o ambiente ao redor tornando assim todo conhecimento contextual, ou seja, não se origina de um fato isolado, mas de uma ou mais interações com o ambiente dentro de uma situação específica tornando grande parte dele vivenciado, sendo assim um conhecimento tácito CAPRA (1996).

No modelo apresentado por MORESI (2000), apresentado na Figura 1, o conhecimento é um dos mais altos no nível hierárquico da informação, abaixo somente da inteligência. O autor define a informação como oportunidade, conhecimento em um contexto relevante que permite atuar com vantagem no ambiente considerado.



**Figura 1 - Os níveis hierárquicos da informação**  
Fonte: Adaptação do autor segundo: MORESI, 2000, p. 18.

## 2.2. Qualidade da informação

A qualidade da informação é um conceito que possui certa dificuldade de definição, por não se ter um consenso entre as suas definições teóricas e operacionais. Porém os autores que abordam sobre o assunto geralmente procuram identificar características para avaliar a informação de modo que atenda a objetivos gerenciais e que possam ser medidas (PAIM et al, 1996). Ao falar ou usar sobre qualquer critério de avaliação a cerca deste assunto, há um caráter incorpóreo, pois é praticamente impossível encontrar um critério que não seja complexo para a medição e que, além disso, seja completamente acurado e que atenda a todas as expectativas e necessidades (SCHWUCHOW, 1990).

### 2.2.1. Medição da qualidade da informação

Ao longo dos anos houveram, várias publicações, tendo como escopo a qualidade da informação e os aspectos relacionados a ela e suas dimensões, ainda a respeito das várias análises em diferentes aspectos, se pode citar Huang, Lee e Wang (1999), Salmeia (1997), Tozer (1997) e SORDI (2008), sendo que este último define a medição da qualidade da informação em dimensões e atributos como pode ser observado no quadro1:

<b>Dimensões da informação</b>	<b>Atributos da informação</b>
Acurácia / veracidade	Nível de acurácia; e método para determinação do nível de acurácia
Atualidade / temporalidade	Data de geração da informação; horário de geração da informação; e intervalo de tempo entre cada nova geração de informação
Disponibilidade	Meio de acesso à informação; horário de disponibilização da informação; e tempo decorrido entre a solicitação e o acesso da informação
Confidencialidade / privacidade	Público-alvo; e predileções informacionais do público-alvo
Existência	Localização do algoritmo para geração da informação; e localização do armazenamento do conteúdo informacional
Abrangência / escopo	Vetores da informação
Integridade	Nível de integridade da informação
Ineditismo / raridade	Disponibilidade de informações idênticas ou similares
Contextualização	Caracterização da informação
Precisão	Nível de precisão da informação
Confiabilidade	Credibilidade da fonte; e credibilidade do conteúdo
Originalidade	Originalidade da informação
Pertinência / agregação de valor	Valor potencial da informação Valor entregue pela informação
Identidade	Nome; sinônimos; e autoria
Audiência	Frequência de acesso; e duração de tempo de acesso

**Quadro 1 - Dimensões e atributos da informação**

Fonte: SORDI (2008).

### 2.2.2 Dimensão e atributos relacionados à informação

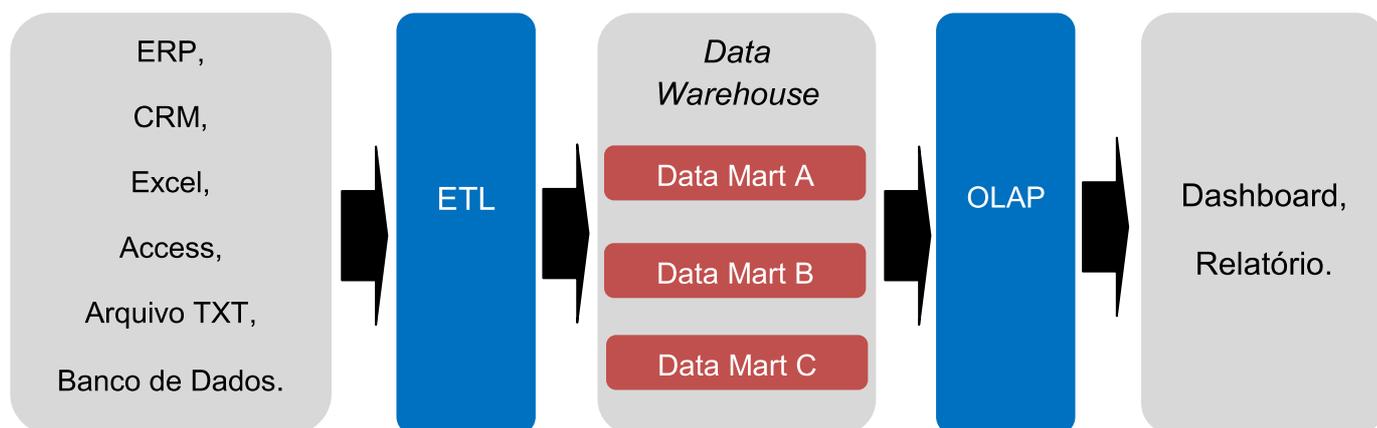
Para avaliar a qualidade da informação, SORDI (2008) apresenta dimensões e atributos da informação, em que as dimensões são as suas características e os atributos são os valores destas propriedades. A respeito dos atributos da informação na busca pela sua qualidade um dos mais importantes são os que dizem a respeito a sua confiabilidade já que uma informação quando não confiável aumenta a desconfiança do gestor, declinando o nível de credibilidade e conseqüentemente á queda da relevância da informação em uma tomada de decisão.

### 2.3 Sistemas de *Business Intelligence*

Necessidades de negócio são os requisitos de informação de um usuário final de um sistema de Business Intelligence (KIMBALL, 2004). Durante o dia a dia de uma empresa há um grande fluxo de entrada e saída de informações e armazenamentos de dados, de modo que se torna dificultoso extrair informações para se enxergar ameaças e oportunidades de negócio para a empresa, sem uma ferramenta de gestão de informação.

O termo Business Intelligence (BI) surgiu na década de 80, pela empresa Gartner Group, (o faz como referência) e é definido como processo de coleta, organização, análise, compartilhamento e monitoramento de informações que ofertam suporte a gestão de negócios (PRIMAK, 2008). Uma solução de Business Intelligence, satisfaz os processos para tratamento de informação proposto por MORESI (2000), no qual visa um sistema de informação para o tratamento de informação em um sistema decisório de uma empresa, onde os dados coletados são trabalhados, filtrados, classificados de modo a obter a informação desejada, com o mínimo de esforço, como demonstrado na Figura 1, em que os dados são processados em informação e analisados pelo usuário para se tornar parte do conhecimento do mesmo.

A Figura 2 demonstra a seqüência de processos em um ambiente de um sistema de Business Intelligence já em execução e seus respectivos componentes.



**Figura 2 - Componentes do ambiente de *Business Intelligence*.**  
 Fonte: Adaptação do autor segundo (BARBIERI, 2001 p.50).

### 2.3.1. ETL

Segundo KIMBALL (2004), o processo de ETL<sup>1</sup> (*Extract, Transform, Load*) cuja tradução significa Extração, Transformação e Carga é um dos processos mais importantes em um sistema de *Business Intelligence*. A função correta de uma ferramenta de ETL e ser um sistema de extração de dados, coletando dados provindos de origens distintas, manter e reforçar a qualidade e consistência dos dados, de modo que possam ser usados em conjunto, possibilitando que os desenvolvedores de aplicativos construam novas aplicações ou façam o uso de suas aplicações já existentes, para mostrar aos usuários finais os dados coletados, deste modo obtendo informações e conseqüentemente tomar as suas decisões.

Ainda segundo KIMBALL (2004), o processo de ETL é o mais longo, em questão de tempo e esforço de um ambiente de BI, consumindo facilmente cerca de 70% dos recursos de um projeto, somando os esforços de implantação e manutenção gastos.

A longa durabilidade de tempo do processo de ETL pode vir a ser uma inconveniência no projeto, pois este processo não é visível ao usuário final, ou

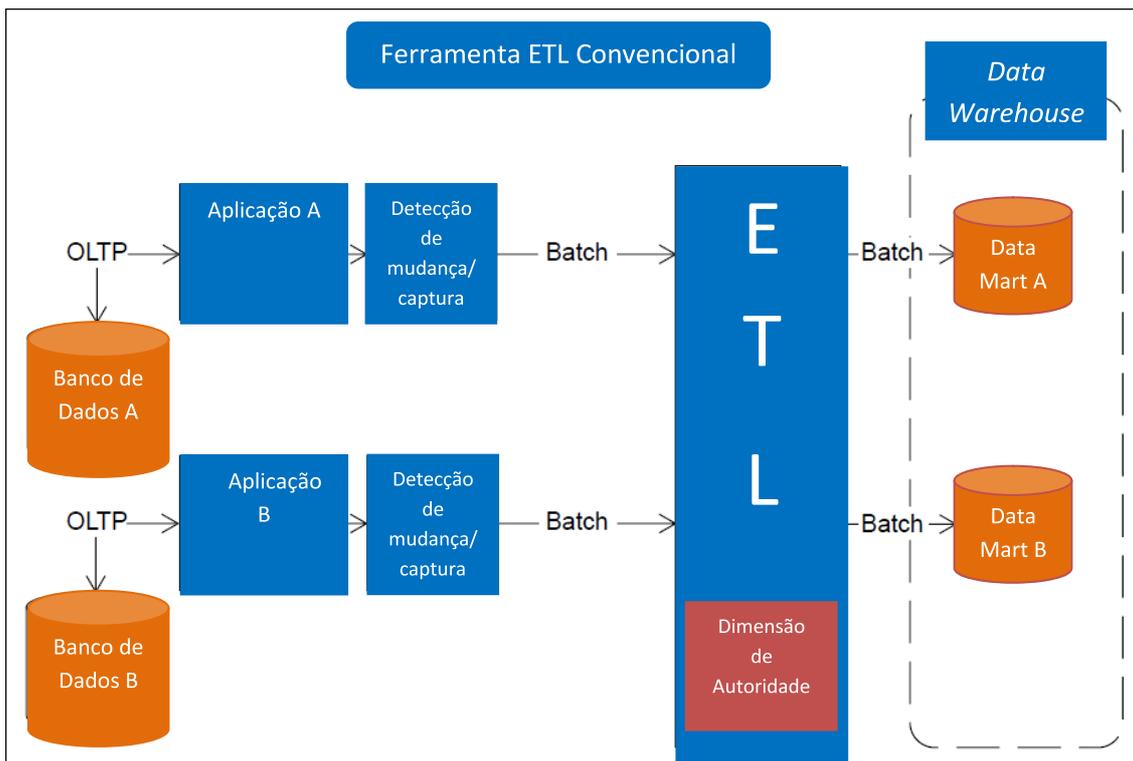
<sup>1</sup> Que pode ser entendido como: Extração como coletar a informação; Transformação como adequação dos dados a um determinado padrão; Carga como envio de informações para a base de dados.

seja, ele não participa desta etapa e dificilmente tem o sentimento de fluência do projeto, o que pode levá-lo a pensar que o projeto está lento ou estagnado (KIMBALL, 2004).

Para KIMBALL (2004), até os dados estarem aptos a serem integrados aos seus respectivos *Data Marts* do *Data Warehouse* e serem manipulados pelas aplicações que os expõem ao usuário final, são realizadas as seguintes etapas com os dados durante o processo de ETL:

- Remoção de erros e correção de falta de dados.
- Fornecimento de medidas documentadas de confiança nos dados.
- Captação do fluxo de dados transacional para a captura.
- Ajuste dos dados de múltiplas fontes para ser utilizado em conjunto.
- Estruturação de dados a serem utilizados por ferramentas de acesso do usuário final. Apresentação dos dados.

KIMBALL (2004), ainda define que, existem dois tipos de ferramentas ETL: O primeiro dos dois tipos são as que têm capacidade de realizar os processos convencionais e podem ser vistos na figura 3. O segundo tipo são as ferramentas de ETL prontas em forma de *Kit de softwares* e as ferramentas feitas através de codificação manual ou própria. A diferença entre ambas é que enquanto a primeira é uma ferramenta já pronta e tem a necessidade de configuração para identificar as fontes de origem e os seus respectivos dados para coleta, à segunda já é uma solução dita como mais caseira, que ocorre quando o profissional cria uma ferramenta que atenda as suas necessidades de coleta. Deste modo, há a possibilidade de utilizar uma ferramenta “pronta” cuja necessidade se torna a de parametrizar e padronizar as suas rotinas somente, o que possui como vantagem a economia de tempo e facilidade para iniciar o processo seu funcionamento, porém como desvantagem o preço maior incluso no projeto para adquirir um *software* com esta finalidade ou criar uma ferramenta totalmente nova, possuindo o benefício de ser extremamente personalizável e sem custos adicionais ao projeto, porém o malefício de gasto maior dos recursos humanos, além do sensível maior consumo de tempo.



**Figura 3–Diagrama de uma ferramenta de ETL convencional**  
 Fonte: Adaptação do autor segundo (KIMBALL, 2004 p.473).

### 2.3.2. Data Warehouse

Resultante da necessidade de negócio para obter informações, (KIMBALL, 2004), *Data Warehouse*, em tradução literal para a língua portuguesa, significa armazém de dados, o que muito bem traduz a sua finalidade, pois seu objetivo é armazenar os dados de distintos sistemas, que são depositados nele pelas ferramentas de ETL.

KIMBALL, (2002) alerta que *Data Warehouse* não é apenas uma mera cópia do sistema operacional de uma organização, levando em consideração que este tipo de sistema é usado para as atividades cotidianas, visto que diversos aspectos que diferenciam os dados armazenados pelo *Data Warehouse* e pelo sistema operacional. Ainda segundo KIMBALL (2002), ambos são os dois locais onde são depositados os dados da organização e podem ter dados em comum.

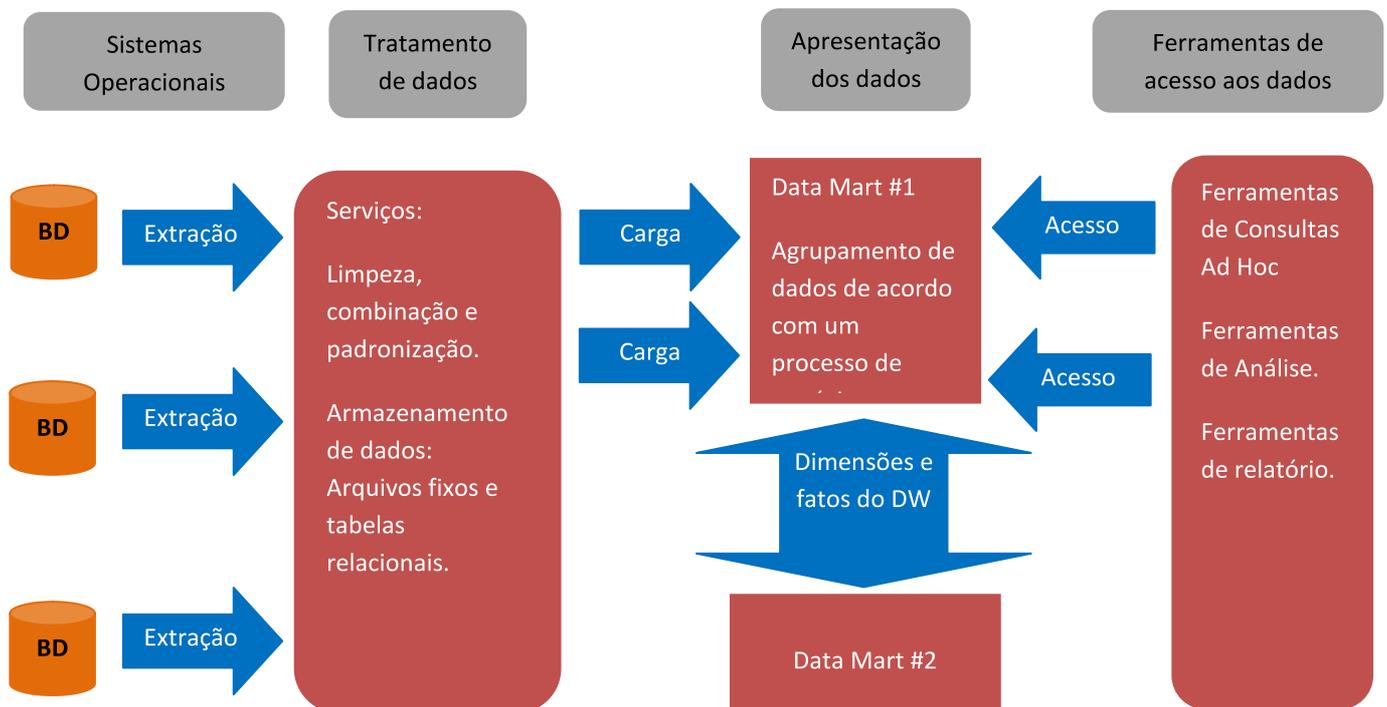
As diferenças entre um sistema operacional e um *Data Warehouse*, são amplas, como por exemplo: o tipo de informação que é armazenada em cada um, enquanto o sistema operacional procura atender os aspectos para que a engrenagens de processos da empresa continuem em funcionamento, o *Data Warehouse*, procura ter contido em si, dados que permitem verificar o desempenho destas engrenagens (KIMBALL, 2002).

Antes de planejar e começar a por em prática um *Data Warehouse*, se deve checar se o mesmo atende a alguns requisitos expostos por (KIMBALL, 2002):

- Ser facilmente acessível. (interface de acesso aos dados intuitivos e de fácil entendimento ao gestor e não somente ao desenvolvedor, de modo que o usuário final consiga manipular os dados para obter a informação que deseja com o mínimo de esforço e tempo de resposta possível).
- Deve apresentar para a organização, informações consistentes. (Os dados contidos no *Data Warehouse* devem ter credibilidade, pois é a reunião de uma variedade de dados coletados em diferentes sistemas, já filtrados, limpos e com a qualidade assegurada através do devido tratamento de manuseamento executado no processo de ETL).
- Devem ser adaptáveis. (Não é devido, ignorar as possibilidades de mudança e suas causas, como condições do negócio, necessidade do usuário, entre outros fatores, todos estão sujeitos á mudança. Então o *Data Warehouse* não deve ser projetado de um modo engessado, que não permita, ou se torne um empecilho para a modificação necessária de acordo com a mudança prevista).
- Protegido, com a segurança da informação assegurada. (Em um *Data Warehouse*, tem dentro de si as informações em forma de joia, ou seja, com dados, que são coletados em pedra bruta já lapidados, que se deve evitar o acesso de pessoas erradas, por isso se faz necessária à segurança de acesso a este tipo de informação).
- Servir de base para tomada de decisão. (Deve conter os dados certos, para ser a fundamentação assertiva para escolha de decisão que o gestor poderá tomar, por que o fator mais importante nos dados que são expostos ao usuário final, é que as decisões só são tomadas após as evidências apresentadas nas consultas feitas ao *Data Warehouse*).

- A aceitação da comunidade de negócio da organização. (Os usuários finais do sistema de suporte a decisão devem abraçar o sistema, isto é aceita-lo e usá-lo. A não utilização do sistema pelos usuários destinados ao uso, seis meses após o treinamento, denunciará a falha no teste de aceitação do usuário, o que é vital para o sucesso final do projeto).

Na figura 4, é possível observar os elementos e processos básicos para a constituição de um *Data Warehouse* segundo KIMBALL (2002).



**Figura 4–Elementos Básicos de um Data Warehouse**  
 Fonte: Adaptação do autor segundo (KIMBALL, 2002 p.7).

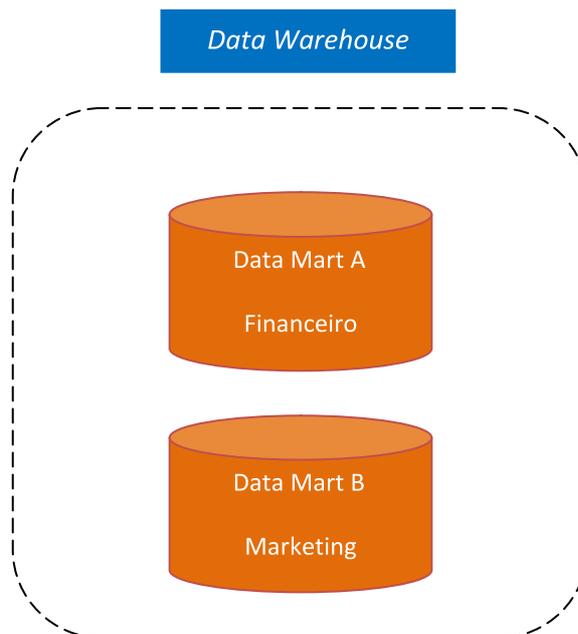
### 2.3.3. *Data Mart*

*Data Mart* é a porção dos dados de um *Data Warehouse* que são separados e organizados para um processo único de negócio, sendo que estes processos podem atravessar as fronteiras de funções organizacionais (KIMBALL, 2002).

KIMBALL (2002) faz uma analogia de que um *Data Mart* está para o *Data Warehouse*, como uma fatia está para um bolo. Nesta similitude, ele procura exemplificar o fato de que o *Data Mart* é uma porção (ou uma fatia) do *Data Warehouse*, no caso é uma parte, um conjunto de dados selecionados para o levantamento de informação, geralmente filtrados e selecionados por processo ou área.

A motivação do uso de *Data Marts* está no fato do longo tempo gasto para a implantação do *Data Warehouse* por completo, além do alto encargo financeiro, por isso, muitos projetos são realizados voltados para algumas áreas ou departamentos da empresa, como *Marketing* e Financeiro, por exemplo. Através do uso de *Data Marts* se obtém um resultado com menor tempo e investimento, além de permitir que o projeto tenha a possibilidade de alcançar outras áreas que irão compor o *Data Warehouse* final (BARBIERI, 2001).

O exemplo de uso dos *Data Marts* pode ser visto na figura 5 onde os dados do *Data Warehouse* são organizados por área da empresa.



**Figura 5–Data Marts em um Data Warehouse**

Fonte: Adaptação do autor segundo (KIMBALL, 2004 p.473).

#### 2.3.4. OLAP

O Termo OLAP, ou *Online Analytic Processing*, foi pela primeira vez citado por E. F. Codd em 1985, que define que OLAP é uma categoria de processamento de banco de dados, que oferece a capacidade de consolidar, visualizar e analisar dados de acordo com múltiplas dimensões. Segundo KIMBALL (2002) é um processo constituído por um conjunto de princípios para fornecer uma estrutura dimensional de suporte à tomada de decisão.

É a ferramenta de OLAP que proporciona ao usuário acessar os dados que estão inseridos dentro do Data Warehouse e seus Data Marts, gerando informação para o gestor por meio de visualizações de relatórios, possibilitando empregar os novos conhecimentos adquiridos sobre o negócio em sua tomada de decisão.

Segundo CODD (1993) uma ferramenta OLAP deve ser vista pelo usuário final como uma planilha com a possibilidade de customização, onde os dados seriam ordenados de acordo com a necessidade desejada, para isso ele define doze regras para uma ferramenta OLAP seguir, que são:

1. Conceito de Visão Multidimensional.

Para os usuários há uma necessidade de visualização de dados de modo multidimensional, para que, por exemplo, os lucros possam ser vistos por região, produto, período de tempo, ou em cenários como orçamento real ou previsão. Modelos de dados multidimensionais permitem a manipulação mais simples e intuitiva dos dados por parte do usuário.

2. Transparência.

O OLAP deve ser parte de uma arquitetura de sistemas abertos, que podem ser incorporados em qualquer lugar desejado pelo usuário, sem afetar negativamente a funcionalidade da ferramenta que esta acolhendo, mas o sistema utilizador não deve ter acesso á fonte da base de dados de onde provêm as informações geradas pela OLAP, tanto os homogêneos quanto heterogêneos.

3. Acessibilidade

A ferramenta deve ser capaz de aplicar a sua própria estrutura lógica de acesso a fontes heterogêneas de dados e realizar as conversões necessárias para apresentar uma vista coerente para o utilizador. A ferramenta (e não o usuário), deve se preocupar com o local de onde vêm os dados físicos.

4. Desempenho de geração de relatório consistente.

O Desempenho da ferramenta de OLAP não deve sofrer nenhuma perda significativa, à medida que as dimensões são aumentadas.

## 5. Arquitetura Cliente/Servidor

O componente servidor da ferramenta de OLAP deve ser inteligente o suficiente para que vários clientes se conectem nele com esforço mínimo. O servidor deve ser capaz de mapear e consolidar os dados entre as bases de dados distintas.

## 6. Dimensionamento genérico

Cada dimensão dos dados deve ser equivalente em sua estrutura e capacidades operacionais.

## 7. Manuseamento de matriz dinâmica esparsa

Estrutura física do servidor OLAP deve ter o tratamento de matriz de forma independente.

## 8. Suporte a múltiplos usuários

A ferramenta deve fornecer consulta simultânea e acesso a atualizações dos dados, além de integridade e segurança.

## 9. Possibilidades de cruzamento de dimensões irrestritas

Fornecer facilidades computacionais com o intuito de permitir a manipulação de cálculo e dados através de qualquer número de dimensões de dados de modo irrestrito.

## 10. Manipulação de dados intuitiva

Prover a manipulação de dados de modo intuitivo, inerente aos processos de consolidação, como aumento e diminuição da granularidade das dimensões de visualização.

## 11. Relatórios flexíveis

O reporte dos relatórios deve apresentar a informação de qualquer maneira que o usuário deseja visualizar.

## 12. Dimensões e níveis de agregação ilimitada

O número de dimensões de dados suportados deve ser para todos os efeitos ilimitados. Cada dimensão genérica deve permitir um número ilimitado de níveis de agregação, definido pelo usuário, em qualquer sentido de compilação de dados.

### 2.4. Modelagem Multidimensional

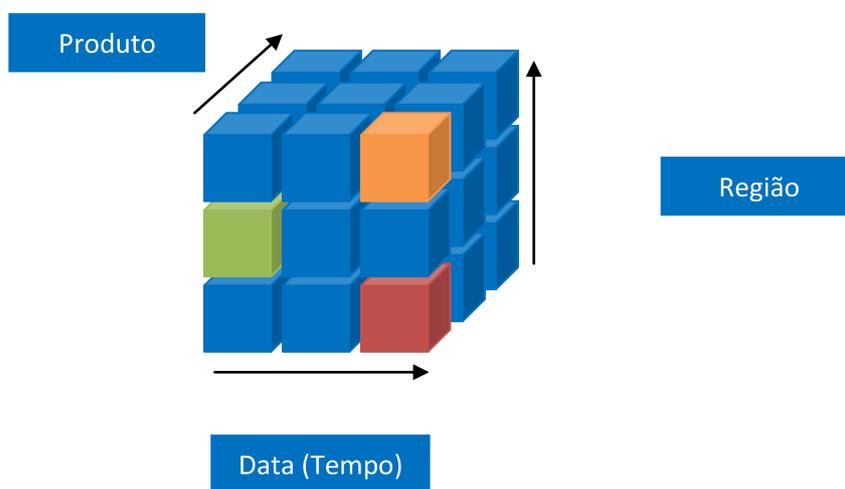
As tecnologias empregadas na ferramenta OLAP permitem manipular os dados inclusos em um *Data Warehouse* através de uma estrutura multidimensional denominada cubo. Os cubos são representações dos dados no *Data Warehouse* e é organizado de maneira a capacitar a execução de consultas rápidas voltadas à obtenção de informação para o apoio a decisão. Os cubos são definidos através de dimensões e medidas (fatos). As dimensões fornecem as informações descritivas, enquanto que as medidas disponibilizam informações quantitativas que se deseja consultar (KIMBALL, 2002).

Os cubos utilizados na modelagem multidimensional possuem suas características e elementos, que pode variar de um para o outro, tal como a modelagem de suas estruturas com o intuito de ao cruzar os dados presentes nas tabelas de dimensão (por isso o termo multidimensional) para se auferir a medida ou fato. Apesar de ser comumente usado em estruturas multidimensionais o conceito de cubo pode ser aplicado em modelagens relacionais (geralmente aplicada a sistemas operacionais da empresa), apesar dos escopos de modelagem se diferir, segundo (KIMBALL, 2002).

Na figura 6 é demonstrado um cubo OLAP em que há o cruzamento de três dimensões (Produto, Região e Tempo), sendo que cada pequeno cubo é uma visualização diferente, uma forma de visão de um fato ao usuário final, com a possibilidade de navegação entre estes cubos.

Além da navegação entre os cubos, é possível também tanto o agrupamento em cubos maiores, quanto à quebra de cubos em fragmentos de cubos menores.

Quando um dos parâmetros<sup>2</sup> de uma ou mais dimensões de um cubo é modificado, seja por navegação entre as datas, passando de um mês para outro, por exemplo, ou por aumento (*Drill Up*) ou diminuição (*Drill Down*) da granularidade de uma dimensão, que compõem o cubo, como por exemplo: ao invés de o agrupamento presente na dimensão região, passar de região de um grupo hipotético de cidades, definido pela organização, passar a ser por estados brasileiros, neste caso há o aumento da granularidade (agrupamento) e aumento do cubo e quando ocorre o processo inverso, tem se a diminuição da granularidade, ou seja, fragmenta-se o cubo maior com os estados em cubos menores com as cidades.



**Figura 6– Representação de um cubo OLAP**  
 Fonte: Adaptação do autor segundo (GAMA, 2009).

Existem modelos para modelagem dimensional, com o intuito de auxiliar na estruturação das tabelas para efetuar a criação dos cubos OLAP.

KIMBALL (2002) cita dois modelos para a efetuação desta tarefa que são: *Star Schema* (esquema estrela) e *Snowflake Schema* (esquema floco de neve).

*Star Schema*, ou esquema estrela é descrito por KIMBALL (2002), como uma representação de um modelo dimensional de um banco de dados relacional em que uma tabela fato possui uma chave composta e esta é unida a várias

<sup>2</sup> Parâmetros no ambiente computacional são entendidos como referências de valores para o funcionamento determinada função ou processo em um software, este valor poder constante ou variável, exemplo de parâmetros: Variável: Selecionar o mês em que se deseja visualizar uma informação. Constante: O Valor de PI em qualquer cálculo.

tabelas de dimensão, cada uma possuindo uma chave primária. A vantagem de usar este tipo de modelagem está relacionada com o desempenho das consultas executadas, visto que há o relacionamento direto das tabelas de dimensão com a tabela fato, porém tem como desvantagem, consumir um maior espaço em disco, um exemplo deste esquema pode ser observado na figura 7.



**Figura 7–Representação de um modelo *Star Schema***

Fonte: Adaptação do autor segundo (KIMBALL (2002) p.22).

Já o *Snowflake Schema* ou esquema floco de neve, se diferencia pelo fato de suas dimensões e as respectivas tabelas referentes obedecerem à ordem de hierarquia, deste modo há dimensões intermediárias entre algumas tabelas, antes de fazer o relacionamento com a tabela fato. A vantagem do uso deste tipo de modelagem é a melhor organização das tabelas, como pode ser visto na figura 8, porém esta vantagem tem como custo uma carga de processamento maior, podendo influenciar no desempenho.



**Figura 8–Representação de um modelo *Snowflake***

Fonte: Adaptação do autor segundo (KIMBALL (2002) p.55).

## 2.5 Estrutura do Projeto

É recomendável o uso de um padrão de etapas para a execução de um projeto, pois deste modo organiza as tarefas como um todo, auxiliando consideravelmente o planejamento e execução do projeto. Para este tipo de aplicação se destaca mundialmente no meio corporativo, o modelo proposto pelo PMI, parte de seu reconhecimento empresarial se deve ao fato de que os seus conceitos podem ser aplicados a praticamente qualquer projeto, independente de sua natureza.

Foi adotada para o projeto a metodologia proposta pelo PMI. O PMI ou *Project Management Institute* (Instituto de Gerência de Projeto) é composto por profissionais correlacionados à área de gestão de projetos, uma das maiores do mundo no gênero, que tem como finalidade formular padrões de gestão, promover a gestão de projetos como profissão e certificando os mesmos entre outras finalidades. Segundo o Instituto de Gerência de Projetos, os grupos de processo são:

- Iniciação (etapa para definir um novo projeto. Há a definição do escopo inicial).
- Planejamento (definir e refinar os objetivos e analisar os melhores caminhos para alcançar esses objetivos propostos pelo projeto).
- Execução (consiste em processos realizados para definir, o que foi desenvolvido e proposto na etapa de Planejamento de forma a cumprir as especificações do projeto).
- Monitoramento e Controle (consiste nos processos necessários para acompanhar, revisar e verificar o desempenho do projeto, visualizar as áreas nas quais necessitam de alteração no plano proposto e iniciar estas mudanças necessárias).

- Encerramento (processos com o intuito de finalizar as atividades do projeto, visando completar formalmente o projeto).

### 3. METODOLOGIA

Este trabalho consiste em um estudo de caso a respeito da credibilidade da informação repassada aos gestores através de indicadores.

Para o início do projeto, foram realizadas entrevistas com o gerente geral e outros gerentes de área de atuação compatível com as áreas abrangentes ao escopo do projeto. Além dos gerentes, foram realizadas entrevistas com membros da organização, que poderiam ser consumidores em potencial das informações que seriam trabalhadas.

Durante as entrevistas para a percepção das necessidades do usuário, foi possível investigar implicitamente o que na visão dos entrevistados (não tinham total confiança) reduziam a credibilidade da informação. Alguns dos motivos mais citados foram: desconhecimento da origem ou a mesma não era totalmente confiável já que em algum momento no passado houve algum transtorno com a fonte.

Após a análise foi detectada a baixa credibilidade, tal aspecto de fragilidade culminou na elaboração de um projeto de desenvolvimento e implementação de uma ferramenta de divulgação de indicadores que abordavam aspectos de redução do consumo de tempo e qualidade produtiva, em uma empresa de telecomunicação, com o intuito de assegurar a confiabilidade da informação, eliminar divergências e assegurar a credibilidade, para isso usando técnicas de Business Intelligence.

No contexto do trabalho, possuía-se a meta qualitativa de informação que era a de que os dados coletados se transformassem em algo confiável perante os gestores, servindo conjuntamente como elementos para a interpretação dos resultados apresentados. Para demonstrar a importância da qualidade da informação, além de justificar o princípio do projeto de modo teórico, foi diagnosticada a sensível necessidade de demonstrar com fundamentos

teóricos o mapeamento das defasagens da informação e como poderiam ser supridas.

Para esta explanação, referenciais teóricos foram consultados para ajudar neste mapeamento da qualidade da informação, após análises de autores, como DAVENPORT (1998), LEITAO e PITASSI (2002) e SORDI (2008), optou-se por utilizar esse último como referencial teórico, pois além de ser um estudo atual em relação à parte dos outros pesquisados, atendeu a necessidade de se mapear as informações e suas características, podendo-se estabelecer um quadro comparativo destes atributos antes e depois da execução do projeto.

Usando a referência teórica de SORDI (2008), as dimensões usadas para medir a informação foram elencadas na seguinte ordem:

- Confiabilidade:

Transparência em relação à origem dos dados e acesso as métricas executadas, além de fácil contato para sanar as dúvidas.

- Integridade:

Verificação dos dados oriundos do contratante, realizando a validação quantitativa e qualitativa.

- Atualidade/Temporalidade:

Com o ideal diário, mas com a atualização no mínimo semanal, dimensão sujeita a disponibilidade da contratante.

- Disponibilidade:

Disponibilização via navegador, garantindo assim a não alteração dos dados publicados e fácil acesso.

Já com ciência das expectativas dos usuários em potencial, entrou-se em contato com os responsáveis pelas bases de dados da empresa contratante da empresa em questão, no caso a organização que contratou os serviços da empresa de telecomunicações em que ocorreu este projeto, já que foi determinado, antes mesmo deste projeto, que a contratante ficaria incumbida do envio e compartilhamento dos dados, que seriam usados como insumo para a geração de indicadores de desempenho previstos em contrato.

### 3.1 DESENVOLVIMENTO

Com a utilização dos grupos de processos presentes na metodologia do Instituto de gerência de projetos (PMI, 2008) houve o seguinte roteiro de ações buscando organizar as etapas do projeto:

#### 3.1.1 Iniciação

Nesta etapa de iniciação, há como alvo definir o escopo do projeto, bem como definir qual seria a ordem da fila de implantação dos indicadores, além de qual área. Durante as reuniões de alinhamento realizadas, o gerente geral definiu a ordem das áreas que iriam ter os seus indicadores publicados. Além disso, delegou quem seria a pessoa responsável por sanar as dúvidas relacionadas á aspectos técnicos dentre outros fatores que abrangessem a regra de negócio, além de ser a pessoa responsável por validar os indicadores gerados e também qual o nível de qualidade perseguido.

#### 3.1.2 Planejamento

Durante a etapa de planejamento, procuraram-se maiores detalhes sobre os requisitos do projeto, usando métodos de investigação, como entrevistas verbais, análise de expectativa e qual a necessidade do uso das informações que seriam fornecidas, possibilitando deste modo o mapear das principais dimensões da qualidade da informação a serem satisfeitas, tanto as que seriam

cruciais de serem solucionadas, tanto as que seriam sensíveis do ponto de vista do usuário bem como qual seria o padrão das decisões a serem tomadas com as informações entregues. Assim podendo se antecipar ao usuário, quando o mesmo solicitar alterações, incrementos posteriores, além de novos indicadores não definidos no escopo inicial do projeto.

### 3.1.3 Execução

Foi posto em prática o que foi previamente planejado para o desenvolvimento do projeto, obedecendo às métricas e diretrizes estabelecidas pelo escopo.

Foram feitos os downloads, instalações e configurações dos softwares, definidos pela etapa anterior, a de planejamento, na máquina do desenvolvedor responsável por prover a aplicação, entre outras configurações de acesso, como a permissão de acesso ao servidor no qual ficaria hospedado o sistema de Business Intelligence, acesso idem ao site FTP da contratante, para buscar as bases disponibilizadas, entre outras configurações, e ajustes.

Imprevistos como configurações de programas foram resolvidos com pesquisa através de *sites* na internet e configurações relacionados à rede da empresa ou acesso especiais foram contornados junto ao setor de T.I. .

Foram feitos acordos, devidamente documentados por e-mail, para atender a todos os requisitos mínimos para possuir os recursos necessários para a codificação. Além do início da codificação das ferramentas para realizar os processos da arquitetura de Business Intelligence.

### 3.1.4 Monitoramento e Controle

Antes é preciso entender que, monitoramento e controle possuem significados diferentes neste grupo de processos. Monitoramento consiste em observar de modo técnico aos acontecimentos do projeto, verificando a fluência dos processos, se esta ocorrendo tudo conforme o planejado. Já o controle é atribuído quando há a necessidade de intervenção em algum processo em

execução, como a alteração dos parâmetros de um processo, como por exemplo, remodelar um método de execução de alguma tarefa, seja para melhorar o andamento do processo, seja para contornar alguma situação antes não prevista (PMI, 2008).

O processo de monitoramento foi aplicado enquanto os protótipos das ferramentas de extração, transformação e carga, eram acionados. Durante a realização de suas funções era verificado se estava realizando as tarefas e procedimentos para qual foi programada fazer, com sucesso e se garantiriam os requisitos da dimensão de confiabilidade da informação segundo SORDI (2008). Este mesmo processo de verificação da credibilidade da informação foi executado posteriormente, durante o desenvolvimento da ferramenta que cumpria a tarefa proposta pelo processo de análise online das informações presentes nos Data Marts.

Já o controle era posto em prática, todas as vezes que alguma das ferramentas não atendia as expectativas, ou não coletavam os dados esperados, ou não retornavam o que era tido como expectativa. Desta forma os responsáveis pela codificação, tanto pela interface de consulta da regra de negócio, analisavam a falha de tal forma a mudar fazer o caminho reverso do fluxo que a ferramenta fazia, sendo assim vislumbravam quais eram os dados coletados finais, o porquê estava retornado àqueles valores e quais atribuições e métricas poderiam ser ajustadas de modo a permitir à correta obtenção dos dados. Deste modo a interface sugeria e o codificador, dentro do possível aplicava as alterações solicitadas.

### 3.1.5 Encerramento

Foi realizada uma reunião de encerramento com a participação das partes envolvidas no projeto, com o objetivo de remover todas as pontas que ainda por algum motivo não observado, poderiam ainda estar soltas. Deste modo, todas as dúvidas e questionamentos, que os envolvidos ainda possuíam foram sanados, procurando evitar algum problema futuro em relação a esta situação.

Após a apresentação formal da ferramenta e a aprovação geral, houve a composição de como seria feita a atualização dos dados, como período de tempo e o responsável por este processo cíclico e igualmente do tratamento da interface junto ao usuário.

### 3.2 Modelagem dos dados

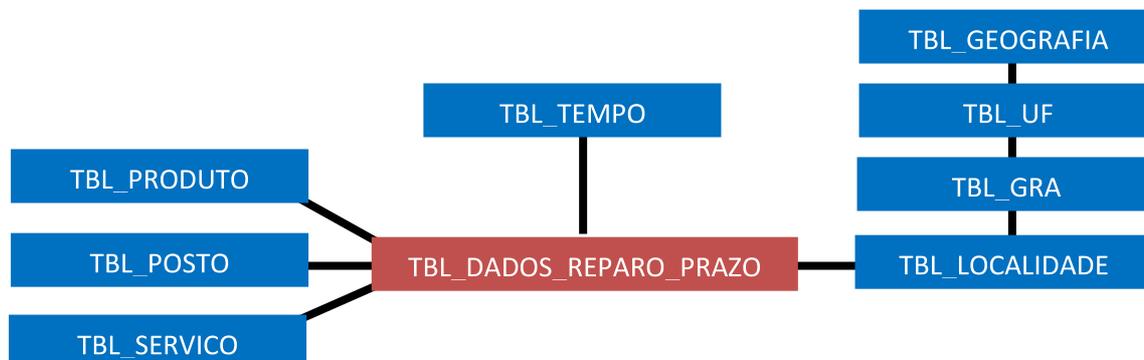
Para a modelagem do Data Mart foi usado o esquema floco de neve, ele foi adotado por que além das características citadas no capítulo 2, como hierarquia e organização, haverá um pequeno número de acessos em relação à quantidade de usuários, a previsão de crescimento dos dados em relação ao custo de processamento esta dentro do aceitável em critérios de desempenho.

Um obstáculo também encontrado no processo de coleta de dados, porém foi possível ser contornado usando o esquema floco de neve, foi o fato que as fontes de onde proviam os dados não eram banco de dados relacionais, eram somente em textos separados por tabulação, ou seja, não se tinha um local por onde aparentemente, se tratavam os relacionamentos entre os dados.

Visto que havia a necessidade de se estabelecer uma escala hierárquica, já que os registros das linhas das bases coletadas para os indicadores são registrados por localidade, o menor grau de agregação de dados existente no escopo da regra de negócio é esta. É a principal referência para os agrupamentos de locais (cidades, municípios, povoados e vilarejos), criando regiões maiores, como GRA (regional geográfica, composta por um conjunto de localidades, que possuem geralmente em seu epicentro uma cidade ou município mais populoso ou desenvolvido da região), estado, e por último, geografia (agrupamento de estados, em grupos específicos) que é a mais alta granularidade.

Esta situação foi encontrada somente nos dados em relação à hierarquia de regiões, além da tomada dessa decisão ter ocorrido em comum acordo durante as reuniões para definição de escopo, a fim de possibilitar uma melhor organização e possibilitar auxiliar no gerenciamento dos dados.

O exemplo do uso de uma modelagem do esquema floco de neve pode ser observado na figura 9.



**Figura 9—Uso do esquema *Snowflake* na diagramação das tabelas**

Fonte: O Autor.

Uma importante observação a ser feita sobre a dimensão tempo e que foi inserida na tabela fato e de que nesta situação segundo KIMBALL (2002), quando uma entidade é usada para identificar uma transação operacional, no caso dados operacionais, são buscados nas fontes provedoras dos dados, como números de controle, ou uma ordem, fatura, entre outros números de filtragem, são chamadas de dimensões degeneradas. Na aplicação do projeto se tem esta situação de dimensão degenerada quando um campo na tabela fato, neste caso o campo referente ao tempo, não provém de uma dimensão.

Houve um cálculo para checar o crescimento do Data Warehouse, para a verificação para que sua expansão não comprometesse o seu desempenho, usando como base as métricas usadas para a geração das tabelas fato, além dos números de registros das extrações preliminares das bases das fontes provedoras de dados. Tendo estas informações, foi possível chegar ao cálculo que uma atualização, teria em média o acréscimo de 2,5 Megabytes no disco rígido, sendo que para o trabalho foram disponibilizados 100 Gigabytes (cada Gigabyte é igual a 1024 Megabytes) de espaço em disco rígido no servidor para a implementação do Data Warehouse.

Os cálculos de consumo em espaço em disco foram assertivos, visto que em novembro de 2011, mês e ano do início da coleta de dados, até meados de novembro de 2012, o tamanho total do banco de dados é de 33,8 Megabytes, acrescentando aproximadamente 35 mil linhas de registro por mês.

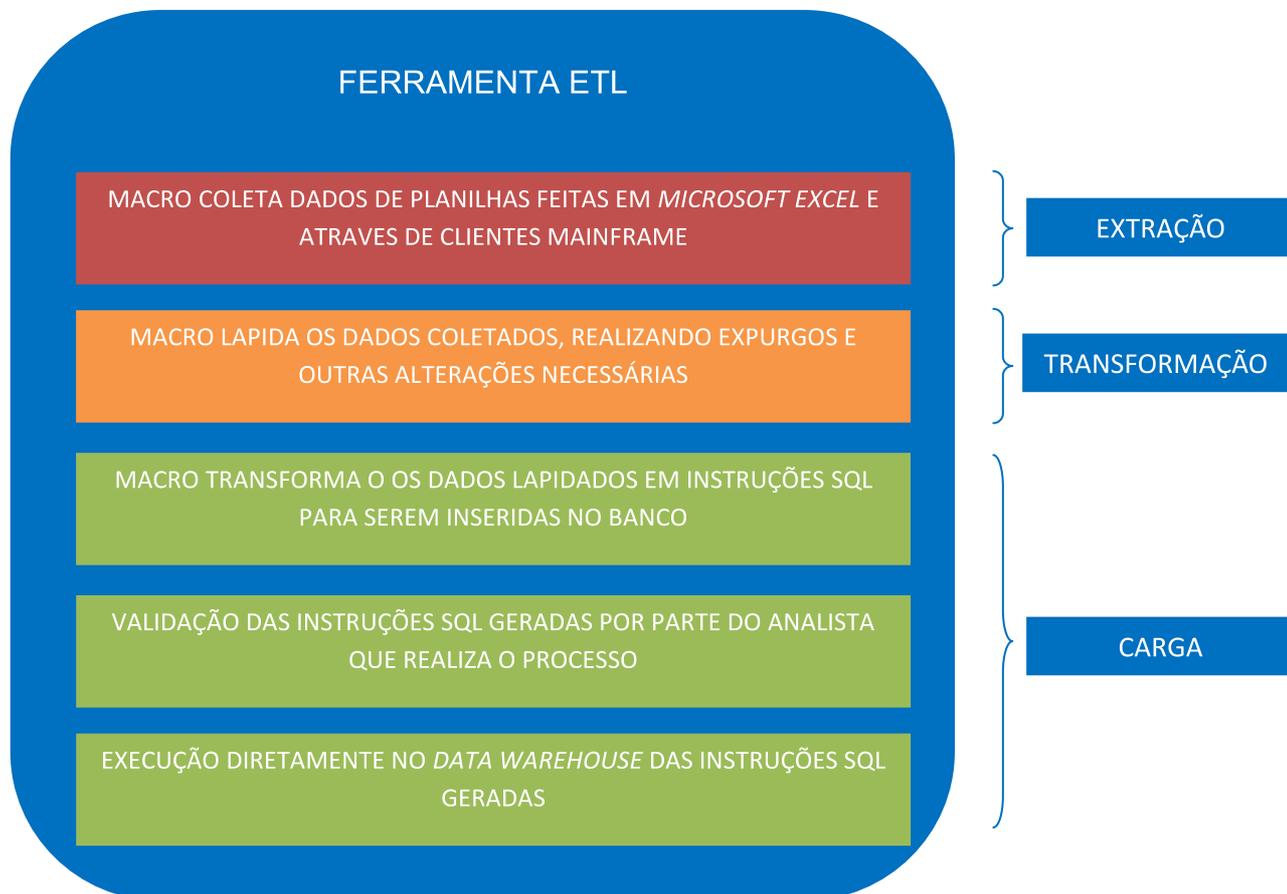
### 3.3. Processo de Extração, Transformação e Carga (ETL).

Segundo KIMBALL (2004) é o processo mais longo e tem a tendência de ser o mais trabalhoso, além da grande probabilidade de consumir cerca de 70% do tempo gasto em um projeto, este evento pode ser produzido com a utilização ou de uma ferramenta já pronta, disponível no mercado, ou através de soluções manuais, aquelas que se aplicam dentro da organização, devido ao fato de um desenvolvedor criar a ferramenta desde o início, com o objetivo de atender as necessidades da coleta de dados envolvendo o seu negócio.

Foi notável, que este processo para efetuar o evento de extração, transformação e carga, fora o que demandou maior esforço e tempo gasto para o projeto. Para concluir esta etapa de ETL, foi desenvolvida uma solução, no software Microsoft Excel, usando neste a linguagem VBA, quando aplicada, a este programa permite a execução de uma macro, processo de automação que permite interagir com outros componentes do computador, simulando um usuário real, assim possibilitando o acesso a sites e sistemas como se fosse um usuário comum, porém com uma velocidade maior.

Para este trabalho, as macros interagiram com planilhas feitas em Microsoft Excel e com programas de clientes de acesso a sistemas Mainframe, para fazer a parte de extração. Após a coleta destas informações, outra macro fica incumbida de filtrar os dados necessários, fazendo assim a parte de transformação e por fim para o processo de carga, é realizado em duas partes: uma terceira macro, tendo como objetivo, criar os comandos em SQL necessários para realizar a carga no banco de dados e a segunda parte consiste na validação dos dados e a execução dos comandos SQL no banco

de dados MySQL, que é o sistema gerenciador de banco de dados presente no Data Warehouse. As ordens dos processos podem ser vista na figura 10.



**Figura 10**–Fluxograma da ferramenta de ETL  
Fonte: O Autor.

### 3.4. Transação analítica de Dados (OLAP)

Após a inserção dos novos dados no Data Warehouse, para atender uma solicitação feita no projeto, um e-mail, é enviado para todos os usuários do banco de dados, anunciando que houve uma nova atualização no banco de dados, para assim terem ciência do mesmo.

A ferramenta de OLAP foi codificada na linguagem de programação PHP, cujo nome é um acrônimo recursivo (palavra ou sigla constituída por letras ou sílabas iniciais de palavras sucessivas de uma expressão, ou pela maioria destas partes) de PHP: Hypertext (Hipertexto) Processor (Processador), é uma

linguagem de script open source (código aberto, livre, gratuito), geralmente utilizada além de especialmente preparada, para o desenvolvimento de aplicações Web, estas acessadas via navegadores (PHP Group, 2001).

Esta solução faz interface do usuário junto à base de dados contida no Data Warehouse, possibilitando um acesso amigável aos dados além de restringir o acesso à base por medida de segurança para evitar, a realização de processos não autorizados à base, já que nesta nenhuma outra operação por parte do usuário é permitida além da visualização, além de proibir o visão indevida dos colaboradores e outros que não deve ter acesso à informação, assim preenchendo grande parte das requisições de como uma ferramenta OLAP deve se portar em um sistema de Business Intelligence segundo CODD (1993).

Na figura11 pode ser vista a interação do usuário com a base de dados, causada pela ferramenta que realiza o processo de OLAP, é mostrada as ações de interações do usuário com a ferramenta, sendo que o cliente da aplicação, por medida de segurança não tem direito a acesso direto a base.

Para atender as suas requisições há uma segunda camada de aplicação, atuando como uma gestora das informações, na qual verifica as credenciais do usuário, como se esta conectado ao sistema e o seu perfil de acesso, permitindo ou negando o acesso a aquela informação.

Por ser acordado em determinadas condições específicas, para a apresentação dos detalhes técnicos do projeto na empresa de telecomunicações deste trabalho, se enquadra em uma destas determinações sendo tida como obrigatória, a não divulgação de qualquer material e reprodução de telas e imagens do sistema, principalmente de relatórios e dados de produção ditos como confidenciais e informações chave para a leitura da produtividade da empresa, sendo assim de forma alguma estas informações, mesmo em caráter de demonstração podem ser exibidas.

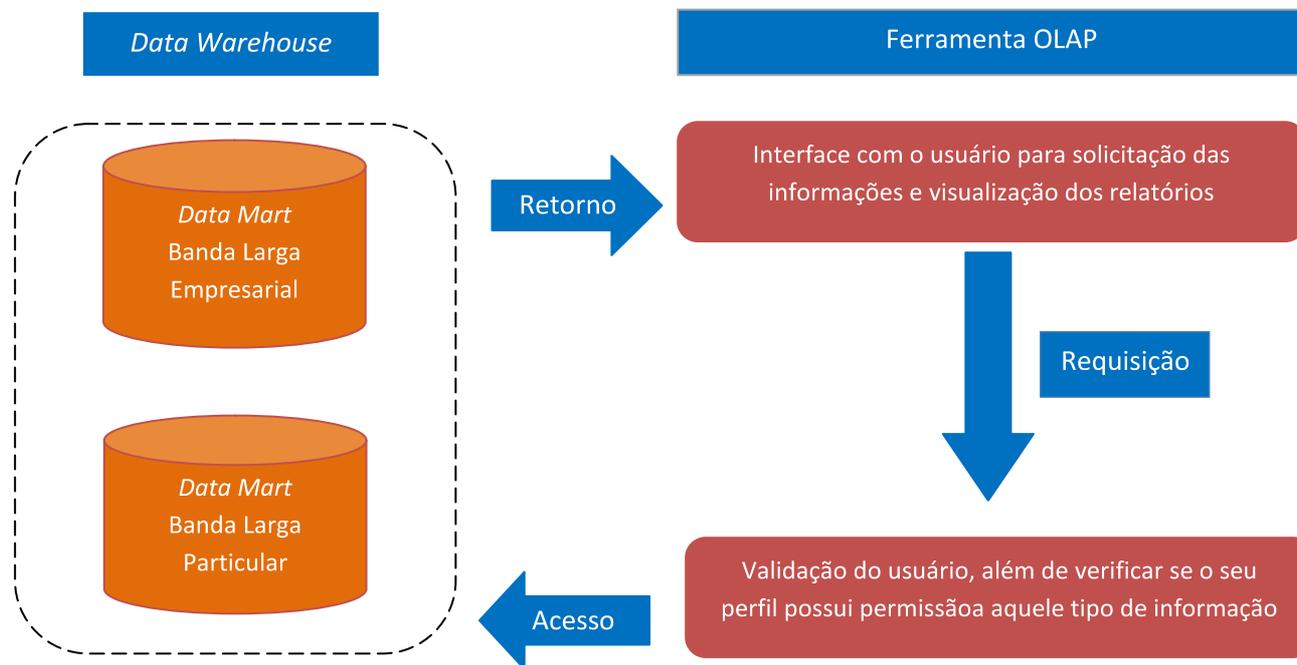


Figura 11 – Fluxo de processos da ferramenta ETL desenvolvida

Ao fim do processo técnico pode ser visto o seguinte relacionamento sobre qual processo da estrutura proposta pela inteligência de negócio preenche os atributos das dimensões escolhidas para validação.

Dimensão	Processo	Descrição
Confiabilidade	ETL	Extração dos dados, aplicação dos expurgos com fundamento nas métricas contratuais.
Integridade	Data Warehouse/Data Mart	Verificação dos dados coletados, analisando possíveis erros.
Atualidade/Temporalidade	Data Warehouse/Data Mart	Os dados após, inseridos no Data Warehouse, já estavam disponíveis para visualização.
Disponibilidade	OLAP	Visualização através de um portal Web.

**Quadro 2 – Tabela com a relação de qual dimensão da informação era atendida em qual processo da arquitetura de *Business Intelligence***

Fonte: o Autor.

#### 4. CONCLUSÃO

O escopo do projeto apresentado neste trabalho foi o de mostrar à resolução do problema da baixa credibilidade da informação na organização através de uma ferramenta de suporte a tomada de decisão, baseada no conceito de *Business Intelligence* usufruindo programas já usados no dia a dia da empresa, entregando informações com a qualificação necessária, preenchendo os atributos de suas propriedades no qual eram esperadas.

Ao longo das etapas contidas no projeto, foi de grande notoriedade o fato de que, com o planejamento assertivo nas etapas iniciais, as ferramentas desenvolvidas na empresa executaram suas tarefas de modo satisfatório, sem sobrecarga em processo algum e que o preenchimento dos atributos de qualidade da informação analisados, foram preenchidos como consequência deste sucesso.

Um dos fatos que contribuiu de modo positivo tanto para a confiança dos gestores nos números apresentados, bem como para a continuidade do projeto, foi à notificação de todos envolvidos no projeto a cada nova visão dos dados, sendo que as visões foram disponibilizadas de modo incremental, sanando o problema já exposto por KIMBALL (2004), que consiste em que por o processo de ETL (extração dos dados) ter o constante rótulo de processo mais duradouro do projeto e não ser visível ao usuário.

A apresentação de cada novo indicador fez com que os usuários tivessem a sensação de andamento do projeto, mostrando que o mesmo não estava estagnado, além que a cada nova visão de dados se abria o espaço para indagações e sugestões, abrindo caminho para uma maior interação entre as partes envolvidas e sanando as questões que ainda ficaram pendentes, como por exemplo: a utilização das métricas na aplicação da regra de negócio. Neste ultimo podendo usufruir do quadro de dimensões da informação para auxiliar nas respostas a serem dadas.

As qualidades da informação dentro das dimensões propostas evoluíram sensivelmente como pode ser visto no modelo no quadro 3 abaixo:

Qualidade Informação		
Dimensão	Valores	
	Antes	Depois
Confiabilidade:	Não existia um único local para extração dos dados, divergências constantes.	Local único definido, métricas transparentes.
Integridade:	Dados poderiam não ser atualizados, ou virem incompletos ou incorretos e a percepção desta era geralmente tardia.	Todo novo arquivo de dados disponibilizado é analisado, caso apresente alguma inconsistência a contratante é acionada.
Atualidade/Temporalidade:	Quando o Gestor solicitava.	Com média atual de 2 vezes por semana.
Disponibilidade:	Via planilha eletrônica enviada por correio eletrônico.	Publicação via sistema que é acessado via navegador de internet, quando ocorre atualização, todos os usuários dos sistemas recebem a notificação por correio eletrônico.

### Quadro 3 – Comparativo de período das dimensões da informação analisadas

Fonte: O Autor.

Grande parte desta agilidade se deve ao esquema Floco de Neve que possibilitou uma melhor organização dos dados e ocasionou de certa forma uma maior facilidade ao transcrever os comandos de SQL para realizar as consultas ao Data Warehouse.

O processo durou o total de cinco meses, de novembro de 2011, a março de 2012, porém houve uma pausa do início de dezembro de 2011, até o início de fevereiro do ano seguinte. Além de que a estrutura da organização da empresa, segundo as classificações do PMBOK (2008) é entendida como matricial fraca, sendo que este tipo de estrutura consiste basicamente, no fato de que a parte operacional da empresa não é orientada a projetos, mas sim no modelo funcional, que em resumo é o modelo mais encontrado nas organizações que respeita o modelo hierárquico através de cargos e departamentos, segundo os critérios estabelecidos pela empresa.

Com a aceitação da parte dos gestores da ferramenta, mesmo que o acesso aos tipos de informações em questão fosse realizado de maneira nova por

parte dos usuários, acostumados a planilhas feitas no Microsoft Excel, com a aplicação de fórmulas, o portal de indicadores de desempenho se tornou a fonte oficial, sanando alguns empecilhos antes encontrados na organização, como por exemplo, não haver um local dito como oficial de provedor desta espécie de informação solicitada.

O sistema para a visualização de indicadores de desempenho atende o que é esperado deste tipo de ferramenta, pois segundo ATKINSON (1998), é esperado que os indicadores preste auxílio na gestão, onde se deve focar nas variáveis críticas para o sucesso da organização e estruturar compensações para quem contribuiu para a eficiência e eficácia dos números.

Na organização, as reuniões de gestão de redirecionamento estratégico têm como auxílio os números apresentados pelos indicadores de desempenho, para a leitura do momento atual da parte operacional da empresa, visto que os indicadores medem o cumprimento ou não dos prazos dos serviços solicitados, além da qualidade destes serviços, métricas estabelecidas pela contratante, rastreando questões que podem afetar a credibilidade da empresa.

Portanto, foi possível a obtenção de um sistema de auxílio à tomada de decisão, obedecendo aos critérios acadêmicos de qualidade da informação recomendados por SORDI (2008) no que diz a respeito da confiabilidade da informação. Isto para a empresa sem um investimento financeiro, em novas aquisições ou contratações de mão de obra, pois as pessoas alocadas já eram funcionários da organização e os recursos tecnológicos usados já eram disponibilizados pela empresa antes do projeto ou foram adquiridos gratuitamente pela internet.

Não houve um retorno financeiro direto, mas sim uma agilidade no processo de análise e tomada de decisão, o que pode gerar um retorno financeiro considerável indiretamente, com o aumento da quantidade e qualidade dos serviços ofertados pela empresa.

Através da execução desse projeto, observou-se que a criação de um sistema de informação, para centralizar informações, pode estabelecer a confiabilidade dos dados e informações consumidas.

## REFERÊNCIAS

AMARAL, Sueli Angélica do; SOUSA, Antônio José Figueiredo de. Qualidade da informação e intuição na tomada de decisão organizacional. *Perspectivas em Ciência da Informação*, Belo Horizonte. Vol. 16, n. 1, mar/2011.

ATKINSON, Anthony. Strategic performance measurement and incentive compensation, *European Management Journal*, Vol. 16, Nº 5, pp552-561, Oct. 1998.

BARBIERI, Carlos. *BI – Business Intelligence: Modelagem e Tecnologia*. 1 ed. Rio de Janeiro: Ed. Axcel Books, 2001. 424p.

BARBOSA, R. R. Acesso e necessidades de informação de profissionais brasileiros: um estudo exploratório.

*Perspectivas em Ciência da Informação*, v. 2, n. 1, p. 5-35, jan./jun.1997.

CAPRA, F. *A teia da vida*. São Paulo: Cultrix 1996.

CODD, E. F., CODD, S. B., SALLEY, C. T. Providing OLAP (On-line Analytical Processing). Codd & Date, Inc. San Jose. Setembro/93 Disponível em: <[http://www.sgpinternational.com/us/products/dataquest/whitepapers/olap\\_wp\\_efcodd.pdf](http://www.sgpinternational.com/us/products/dataquest/whitepapers/olap_wp_efcodd.pdf)>. Acesso em: 18 nov. 2012.

CRUVINEL, José Humberto. Análise do ambiente informacional Prodemge com foco na criação de um portal corporativo. *Revista Fonte Tecnologia da Informação na Gestão Pública*, Belo Horizonte, Ano sete Volume 10.

DAVENPORT, T. H. *Ecologia da Informação: por que só a tecnologia não basta para o sucesso na era da informação*. 3. ed. São Paulo: Futura, 2000. 316 p.

DEZ PASSOS PARA A IMPLANTAÇÃO DE PROJETOS DE *BUSINESS INTELLIGENCE* – Disponível em: 26 de nov. 2012

DEMPSEY, S. et al. The use of strategic performance variable as leading indicators in financial analysts' forecasts. Disponível em: <[http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2346](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2346)> Acesso em: 28 jun. 2014.

ECCLES, R. G.; MAVRINAC, S. Improving the corporate disclosure process. *Sloan Management Review*, v. 36, 1995. Disponível em: <<http://sloanreview.mit.edu/article/improving-the-corporate-disclosure-process/>> Acesso em: 28 jun. 2014.

GAMA, João. Do OLTP ao OLAP. Disponível em: <http://www.liacc.up.pt/~jgama/Bdc/olap.pdf>. Acesso em: 25 de abr. de 2012.

GONÇALVES, G.J. *Sistemas de informação automatizados: uma análise crítica sobre a sua eficácia*. Dissertação (Mestrado) – Instituto de Administração e Gerência, Pontifícia Universidade Católica do Rio de Janeiro, 1996.

LEITAO, Sergio Proença; PITASSI, Claudio. TECNOLOGIA DE INFORMAÇÃO E MUDANÇA: UMA ABORDAGEM CRÍTICA. *RAE - Revista de Administração de Empresas*, São Paulo, v.42, n.2, p. 77-87, Abr./Jun. 2002.

<http://cio.uol.com.br/gestao/2011/01/12/dez-passos-para-a-implantacao-de-projetos-de-business-intelligence/> - Acessado em: 26 nov. 2012.

HUGE, E.C.: "Measuring and rewarding performance." In: HUGE, E.C (ed.) *Total quality – an executive's guide for the 1990s*. Homewood, Illinois, Business One Irwin, 1990

KIMBALL, Ralph. *The Data Warehouse Toolkit: Second Edition*, CAMPUS, 2002.

KIMBALL, Ralph. *The Data Warehouse ETL Toolkit*, CAMPUS, 2004

MORESI, Eduardo Amadeu Dutra. Delineando o valor do sistema de informação de uma organização. *Ciência da Informação*, Brasília, v. 29, n. 1, p. 12-24, jan./abr. 2000. Disponível em: <<http://www.scielo.br/pdf/ci/v29n1/v29n1a2.pdf>>. Acesso em: 02 de novembro 2012.

OLIVEIRA, Marlene de. *Origens e Evolução da Ciência da Informação*. Belo Horizonte: ECI/UFMG, 2004. Texto didático.

O que é PHP? – Disponível em: [http://php.net/manual/pt\\_BR/intro-whatism.php](http://php.net/manual/pt_BR/intro-whatism.php) - Acessado em: 27 Jun. 2014.

PACE, Eduardo Sérgio Ulrich.; BASSO, Leonardo Fernando Cruz.; SILVA, Marcos Alessandro Da.; Indicadores de desempenho como direcionadores de valor. *Revista de Administração Contemporânea*, Curitiba, v. 7, n. 1, jan./abr. 2003. Disponível em: <<http://www.scielo.br/pdf/ci/v29n1/v29n1a2.pdf>>. Acesso em: 27 de Junho 2014.

PAIM, I.; NEHMY, R. M. Q.; GUIMARÃES, C. G. Problematização do conceito “Qualidade” da informação. *Perspectivas em Ciência da Informação*, Belo Horizonte, v.1, n.1, p. 111-119, jan./jun. 1996.

PRIMAK, Fábio Vinícius da Silva. *Decisões com B.I. (Business Intelligence)*. 1 ed. Rio de Janeiro: Ed. Ciência Moderna Ltda., 2008. 152p.

PMBOK: Um Guia do Conjunto de Conhecimentos em Gerenciamento de Projetos. 4. Ed. Project Management Institute; Four Campus Boulevard; Newton Square; Palo Alto, USA. 2008.

RANGEL, Alexandre. *MYSQL: projeto modelagem e desenvolvimento de banco de dados*, ALTA BOOKS.

SANTOS, J. L. et al. Ativos intangíveis: fonte de vantagem competitiva, v. 6, n. 10, Porto Alegre, Contexto, 2006.

SCHWUCHOW, W, Problems in evaluating the quality of information services. In: WORMELL, I. (Org.) *Information quality: definitions and dimensions*. London: Taylor, 1990. P. 69-72.

SETZER, Valdemar. Dado, informação, conhecimento e competência, Depto. de Ciência da Computação da USP, (Publicado no jornal do Grupo Folha Educação No. 27 out./nov. 2004, pp. 6 e 7).

SOLIGO, VALDECIR. Indicadores: Conceito e complexidade do mensurar em estudos de fenômenos sociais. *Est. Aval. Educ.*, São Paulo, v. 23, n. 52, p. 12-25, mai./ago. 2012

SORDI, José Osvaldo de. *Administração da Informação: fundamentos e práticas para uma nova gestão do conhecimento*. São Paulo: Saraiva 2008.

SORDI, José Osvaldo de; MEIRELES, Manuel; GRIJO, Rogério Nahas. Gestão da qualidade da informação da informação no contexto das organizações: percepções a partir do experimento de análise da confiabilidade dos jornais eletrônicos. *Perspectivas em Ciência da Informação*, v.13, n.2, p.168-195, maio/ago. 2008.

TOFLER, Alvim. *A terceira onda: a morte do industrialismo e o nascimento de uma nova civilização*. 26. ed. Rio de Janeiro: Record, 2001. 491p.

URDANETA, I. P. *Gestión de la inteligencia, aprendizaje tecnológico y modernización del trabajo informacional: retos y oportunidades*. Caracas: Universidad Simón Bolívar, 1992.

## GLOSSÁRIO

“*Batch* - Arquivo de lote - geralmente possui a extensão de arquivo”. *bat*” são arquivos utilizados para automatizar tarefas. É apenas um conjunto de tarefas realizadas de modo autônomo em um computador ou rede, através de um script. Sua utilização é geralmente aplicada quando há necessidade de execução de uma ou mais tarefas que podem ser executadas de modo constante levando em consideração dia e hora e que não precisam de intervenção humana.

*Business Intelligence* ou *B.I.* – Arquitetura para tratamento de dados cuja extração geralmente ocorre de mais de uma fonte diferente, com o intuito compilar os mesmos e prover informações aos gestores de modo que auxiliem o mesmo no processo de tomada de decisão.

CRM – Customer Relationship Management ou Gestão de relacionamento com o cliente define todas as ferramentas que automatizam o contato com o cliente de modo a facilitar o mesmo.

Data Mart – “Mercado” de dados, são segmentos de dados dentro de um Data Warehouse são como “módulos” cujo conteúdo de dados são agrupações, seja por segmento ou área de interesse como financeiro, marketing entre outros.

*Data Warehouse* – Armazém de dados em tradução literal é o local onde são depositados os dados oriundos das fontes que alimentam a ferramenta de *B.I.* , podem ser armazenados em um banco maior e único ou pode ser segmentado por mercados (*Data Marts*) que são bancos menores para facilitar a organização e futuras consultas.

ERP - *Enterprise Resource Planning*, ou sistema integrado de gestão empresarial é um programa de computador, dividido por módulos que tem como objetivo armazenar e servir como interface para visualizar e manipular os dados e processos de uma empresa tudo em um único programa, em que cada módulo fica responsável por uma perspectiva funcional específica (Departamento De Pessoal, Contabilidade, Vendas, Compras, são exemplos).

FTP - *File Transfer Protocol* – Sua tradução literal é Protocolo de Transferência de Arquivos, é um protocolo de comunicação em rede que permite uma forma prática e rápida de transferir arquivos, tanto em rede internas (*Ethernet*), quanto em rede externas (*Internet*), pode ser acessado através de navegadores, ou clientes específicos para essa finalidade.

*Mainframe* - Computador de grande porte, dedicado normalmente ao processamento de um volume grande de informações. Podem possuir um sistema, geralmente operacional, que o acesso é remoto e que só pode ser acessados através de clientes (programas de computador) com esta finalidade.

*Microsoft Access* – Software de base de dados possui a capacidade de armazenar dados em tabelas e criar/gravar consultas, possui uma interface gráfica, que permite além de ser usada para a manipulação dos dados, possui a opção de construir formulários e criar visualizações personalizadas dos dados, também possui suporte a linguagem *VBA*.

*Microsoft Excel* – Software de planilha eletrônica, com diversas aplicabilidades de uso, constitui-se de uma interface gráfica composta por células dispostas em forma de matriz por linhas e colunas, possui também formulas que podem ser inseridas nas células com o intuito de automatizar determinados cálculos ou verificações, além de possuir a opção do acréscimo de código *VBA* para automação de tarefas entre outras possibilidades.

*MySQL*– Sistema Gerenciador de Banco de Dados ou SGBD consiste em um conjunto de programas divididos em módulos sendo que cada um provê um serviço, com o objetivo geral de gerenciamento dos dados seja para armazenamento, segurança, consulta entre outros.

*SQL - Structured Query Language* ou linguagem estruturada de dados. É uma linguagem de pesquisa declarativa, sendo padrão na maioria dos bancos de dados cujo objetivo é o de pesquisar os dados, podendo aplicar diversos filtros para retornar os dados com a informação e no modelo necessário.

*Stakeholder* - publico de interesse, termo largamente usado no que se refere à gestão de projetos, descreve a pessoa ou grupo de pessoas que possui interesse no projeto geralmente estratégico, seja por ser diretamente afetado ou ser um investidor do mesmo.

*VBA – Visual Basic for Applications*, implementação da linguagem de programação *Visual Basic* nos programas da Microsoft pertencentes à suíte de programas *Office* é conhecida por sua flexibilidade e é muito usada para a automação de tarefas repetitivas.