

UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS APLICADAS A EDUCAÇÃO
BACHARELADO EM SISTEMAS DE INFORMAÇÃO

**UMA PROPOSTA PARA APLICAÇÃO DE DATA WAREHOUSE NA
SUPERINTENDÊNCIA DE TECNOLOGIA DA INFORMAÇÃO DA UFPB**

GIORDANIA GISELLE GERMINIANO DA SILVA
Orientador: Profa. Ms. Renata Viegas de Figueiredo

RIO TINTO - PB
2014

GIORDANIA GISELLE GERMINIANO DA SILVA

**UMA PROPOSTA PARA APLICAÇÃO DE DATA WAREHOUSE NA
SUPERINTENDÊNCIA DE TECNOLOGIA DA INFORMAÇÃO DA UFPB**

Monografia apresentada para obtenção do título de Bacharel à banca examinadora no Curso de Bacharelado em Sistemas de Informação do Centro de Ciências Aplicadas e Educação (CCAIE), Campus IV da Universidade Federal da Paraíba.
Orientador: Profa. Ms. Renata Viegas de Figueiredo.

RIO TINTO - PB
2014

S586p Silva, Giordania Giselle Germiniano da.

1.1.1.1 Uma proposta para aplicação de *Data Warehouse* na Superintendência de Tecnologia da Informação da UFPB. / Giordania Giselle Germiniano da Silva. – Rio Tinto: [s.n.], 2014.

64 f. : il. –

Orientadora: Profa. Ms. Renata Viegas de Figueiredo.
Monografia (Graduação) – UFPB/CCAIE.

1. Tomada de decisão - sistemas de informação. 2. *Data Warehouse* – computação. 3. *Business Intelligence* - Administração.

GIORDANIA GISELLE GERMINIANO DA SILVA

UMA PROPOSTA PARA APLICAÇÃO DE DATA WAREHOUSE NA
SUPERINTENDÊNCIA DE TECNOLOGIA DA INFORMAÇÃO DA UFPB

Trabalho de Conclusão de Curso submetido ao Curso de Bacharelado em Sistemas de Informação da Universidade Federal da Paraíba, Campus IV, como parte dos requisitos necessários para obtenção do grau de BACHAREL EM SISTEMAS DE INFORMAÇÃO.

Assinatura do autor: _____

APROVADO POR:

Orientador: Profa. Ms. Renata Viegas de Figueiredo
Universidade Federal da Paraíba – Campus IV

Prof. Msc. José Jorge Lima Dias Junior
Universidade Federal da Paraíba – Campus IV

Prof. Msc. Rafael Marrocos Magalhães
Universidade Federal da Paraíba – Campus IV

RIO TINTO - PB
2014

“Se tu o desejas, podes voar, só tens que confiar muito em ti.”

Steve Jobs

AGRADECIMENTOS

Primeiramente a Deus, por ter me dado saúde, Fé, e forças durante todo esse tempo para chegar ao fim de mais uma etapa em minha vida.

A minha mãe, por todo o apoio e todo amor de sempre, fazendo de tudo por mim e para que eu seja sempre uma pessoa melhor, com princípios, e firme em meus objetivos. Tudo que sou hoje devo a você, mãe. Obrigada por tudo, te amo!

Ao meu pai (*in memorian*) que sempre me incentivou nos estudos e acompanhou o início da minha graduação, me motivando e me ajudando no que lhe era possível para que eu seguisse firme e nunca desistisse.

Aos meus irmãos, em especial a Fábio por sempre estar disposto a me ajudar em tudo o que for preciso, e ter sido sempre como um pai pra mim.

Aos demais familiares, especialmente aos meus avós paternos, Angelita e Manoel, por sempre me apoiar com suas palavras de incentivo, e a minha prima Yalle, por se dispor a me ajudar e tornar as coisas menos difíceis para mim.

Ao meu namorado Fernando Carvalho que é muito mais que namorado, é parceiro, amigo, e me deu toda a força para que eu não desistisse de seguir em frente. Obrigada pela compreensão, por acreditar em mim, e por estar sempre disposto a me ajudar em tudo o que lhe é possível.

A minha orientadora Ms. Renata Viegas de Figueiredo, pelos ensinamentos, orientações, contribuições para o meu trabalho, compreensão, apoio, e incentivo para que eu continuasse firme nos momentos em que estava difícil.

Aos professores do curso de Sistemas de Informação, que através de seus conhecimentos contribuíram para a minha formação acadêmica. Em especial ao Prof. Msc. Jorge Dias, obrigada por além de me passar seus conhecimentos, ter me apoiado e incentivado a chegar até aqui.

Aos meus amigos, que torcem por mim e acreditam que sou capaz de alcançar meus objetivos. Em especial a Igor, Larissa, e Tainar, obrigada pelo apoio e torcida para que tudo desse certo.

Aos meus colegas da universidade que fizeram com que a caminhada se tornasse mais leve durante todos esses anos. Em especial a Simone, Diego, Lilian, e Danilo, que além de colegas se tornaram meus amigos, parceiros, e compartilharam comigo muitos momentos. Obrigada pelo companheirismo, apoio, torcida e parceria de sempre.

Aos funcionários da Superintendência de Tecnologia da Informação, obrigada pelo apoio para com o meu trabalho, e pelo fornecimento dos dados necessários para a realização do mesmo. Em especial a William, muito obrigada pelo apoio e disposição a me ajudar no desenvolvimento do meu trabalho.

RESUMO

Business Intelligence é uma importante prática para o processo de tomada de decisões nas empresas. Suas tecnologias, como por exemplo, *Data Warehouse* e *On-line Analytical Processing* permitem facilitar as análises das informações que são realizadas por gestores dentro das organizações. Em uma universidade, por exemplo, sistemas armazenam uma grande quantidade de dados em suas bases, e consultas à estas bases podem levar muito tempo para serem carregadas, devido à quantidade de dados a serem filtrados. Neste caso, dispor de tecnologias de *Business Intelligence* que torne essa filtragem mais fácil e rápida e que gere conhecimento de uma forma analítica e de fácil compreensão, é uma grande vantagem para os gestores em seus processos de tomada de decisões. Este trabalho é uma proposta de aplicação da tecnologia *Data Warehouse* na Superintendência de Tecnologia de Informação da Universidade Federal da Paraíba, com o objetivo de auxiliar os gestores nos processos decisórios dentro da instituição. Na criação do *Data Warehouse* serão aplicados os conceitos de *Business Intelligence*, e os resultados serão analisados com base nos conceitos de *On-line Analytical Processing*.

Palavras chave: *Business Intelligence, Data Warehouse, On-line Analytical Processing*.

ABSTRACT

Business Intelligence is an important practice for the process of decision making in business. Its technologies, such as Data Warehouse and On-line Analytical Processing facilitating the analysis of information that are performed by managers within organizations. In a university, for example, systems that store a large amount of data in their databases, and queries to these databases can take a long time to load due to the amount of data to be filtered. In this case, have Business Intelligence technologies that make the filtering and generation of knowledge easier and faster in an analytical way with a easy understanding rate. It is a great advantage to managers in their decision making processes. This work is a proposal of application of Data Warehouse Technology in Information Technology Oversight of the Federal University of Paraiba, aiming to assist managers in decision-making processes within the institution. In creating the Data Warehouse Business Intelligence concepts will be applied, and the results will be analyzed based on the concepts of Online Analytical Processing.

Keywords: Business Intelligence, Data Warehouse, On-line Analytical Processing.

LISTA DE FIGURAS

| | |
|--|----|
| Figura 1.- Arquitetura de BI | 5 |
| Figura 2.- Arquitetura básica de um DW | 8 |
| Figura 3.- Arquitetura <i>Top - Down</i> | 10 |
| Figura 4.- Arquitetura <i>Bottom - up</i> | 11 |
| Figura 5.- Cubo de Informações | 13 |
| Figura 6.- Exemplo de fatos e dimensões em um processo de vendas..... | 15 |
| Figura 7.- <i>Star Schema</i> ou Modelo Estrela..... | 16 |
| Figura 8.- Modelo <i>Snow Flake</i> ou Floco de Neve..... | 17 |
| Figura 9.- Exemplo de <i>Drill Down</i> no Cubo de Informações | 19 |
| Figura 10.- Exemplo de <i>Drill Up/Roll Up</i> no Cubo de Informações | 20 |
| Figura 11.- Exemplo de <i>Drill Across</i> | 20 |
| Figura 12.- Exemplo de <i>Drill Through</i> | 21 |
| Figura 13.- Exemplo de <i>Slice</i> no Cubo de Informações | 22 |
| Figura 14.- Exemplo de <i>Dice</i> no Cubo de Informações | 22 |
| Figura 15 – Exemplo de <i>Pivoting</i> em tabelas de Dimensão | 23 |
| Figura 16 – Extração dos dados da tabela Alunos no <i>QlikView</i> | 28 |
| Figura 17 – Transformação dos dados na tabela Matrículas | 29 |
| Figura 18 – Criação da Tabela de Fatos | 30 |
| Figura 19 – Modelo dimensional Floco de Neve gerado pelo <i>QlikView</i> | 31 |
| Figura 20 – Processo de ETL e criação do DW através dos dados da base SCA..... | 32 |
| Figura 21 – Visão geral da situação dos alunos matriculados nas disciplinas do curso de SI . | 33 |
| Figura 22 – Índice de reprovação na disciplina Introdução a Programação durante os anos de 2008 a 2013 | 34 |
| Figura 23 – Índice de reprovação na disciplina Cálculo Diferencial e Integral durante os anos de 2007 a 2013..... | 35 |
| Figura 24 – Índice de reprovação da disciplina Introdução a Programação no ano de 2013 ... | 36 |
| Figura 25 – Quantidade de alunos matriculados nas disciplinas no ano de 2013 | 37 |
| Figura 26 – Visão geral dos alunos evadidos do curso entre os períodos 2008.1 a 2013.2 | 38 |
| Figura 27 – Quantidade de alunos evadidos no período 2010.2..... | 39 |
| Figura 28 – Quantidade de alunos evadidos no período 2011.1..... | 40 |
| Figura 29 – Visão geral dos alunos evadidos do curso de SI entre os períodos 2008.1 a 2013.2..... | 41 |

LISTA DE TABELAS

| | |
|--|----|
| Tabela 1.- Informações sobre as tabelas do SCA | 26 |
|--|----|

LISTA DE SIGLAS

| | |
|-------|---|
| BD | Banco de Dados |
| BI | <i>Business Intelligence</i> |
| DM | <i>Data Mart</i> |
| DOLAP | <i>Desktop On-line Analytical Processing</i> |
| DW | <i>Data Warehouse</i> |
| ETL | <i>Extract, Transform and Load</i> |
| HOLAP | <i>Hybrid On-line Analytical Processing</i> |
| MOLAP | <i>Multidimensional On-line Analytical Processing</i> |
| OLAP | <i>On-line Analytical Processing</i> |
| OLTP | <i>On-line Transaction Processing</i> |
| ROLAP | <i>Relational On-line Analytical Processing</i> |
| SCA | Sistema de Controle Acadêmico |
| SGBD | Sistema Gerenciador de Banco de Dados |
| SI | Sistemas de Informação |
| SQL | <i>Structured Query Language</i> |
| STI | Superintendência de Tecnologia da Informação |
| UFPB | Universidade Federal da Paraíba |

SUMÁRIO

| | |
|--|-------------|
| RESUMO | VIII |
| ABSTRACT | IX |
| LISTA DE FIGURAS | X |
| LISTA DE TABELAS..... | XI |
| LISTA DE SIGLAS..... | XII |
| 1 INTRODUÇÃO | 1 |
| 1.1 MOTIVAÇÃO..... | 1 |
| 1.2 OBJETIVOS | 2 |
| 1.2.1 Objetivo..... | 2 |
| 1.2.2 Objetivos Específicos..... | 3 |
| 1.3 ESTRUTURA DO TRABALHO | 3 |
| 2 FUNDAMENTAÇÃO TEÓRICA | 4 |
| 2.1 BUSINESS INTELLIGENCE..... | 4 |
| 2.1.1 Contextualização | 4 |
| 2.1.2 Ambiente de BI..... | 5 |
| 2.1.3 Extração, Transformação, e Carga | 6 |
| 2.2 DATA WAREHOUSE | 6 |
| 2.2.1 Características | 7 |
| 2.2.2 Arquitetura | 8 |
| 2.2.2.1 Arquitetura Top- Down | 9 |
| 2.2.2.2 Arquitetura Bottom-up..... | 10 |
| 2.2.3 Modelo de Dados | 11 |
| 2.2.3.1 Modelo Relacional..... | 11 |
| 2.2.3.2 Modelo Dimensional | 12 |
| 2.2.3.2.1 Fatos e Dimensões..... | 13 |
| 2.2.3.2.1.1 Star Schema(Modelo Estrela)..... | 16 |
| 2.2.3.2.1.2 Snow Flake(Floco de Neve | 16 |
| 2.2.4 On-le Analytical Processing..... | 17 |
| 2.2.4.1 Arquiteturas OLAP..... | 18 |
| 2.2.4.2 Operações OLAP | 18 |

| | | |
|----------|--|-----------|
| 2.3 | CONSIDERAÇÕES FINAIS DO CAPÍTULO | 23 |
| 3 | ELABORAÇÃO DO PROTÓTIPO DE UM DATA WAREHOUSE..... | 24 |
| 3.1 | A EMPRESA..... | 24 |
| 3.2 | PROBLEMÁTICA..... | 24 |
| 3.3 | APLICANDO DATA WAREHOUSE NA STI..... | 25 |
| 3.3.1 | Banco de Dados do Controle Acadêmico..... | 25 |
| 3.3.2 | QlikView..... | 26 |
| 3.3.3 | Processo ETL..... | 27 |
| 3.3.4 | Construindo cenários..... | 32 |
| 3.4 | CONSIDERAÇÕES FINAIS DO CAPÍTULO..... | 41 |
| 4 | CONCLUSÃO | 42 |
| | REFERÊNCIAS BIBLIOGRÁFICAS | 44 |
| | APÊNDICE..... | 47 |
| | ANEXOS | 50 |

1 INTRODUÇÃO

Este capítulo apresenta a parte introdutória envolvida no trabalho. Aqui serão descritos os seguintes tópicos: motivação, justificativa, objetivos (gerais e específicos), e a estrutura do trabalho.

1.1 MOTIVAÇÃO

Dados geram informações, informações geram conhecimento. Partindo deste princípio, nota-se que há um processo para se chegar ao conhecimento. Dados são considerados elementos que por si só não podem transmitir uma mensagem, para que isto ocorra, é necessário que eles sejam coletados, armazenados, e tratados e, a partir disto, podem ser considerados informações. Informação é um conjunto de dados que foram tratados e interpretados para serem compreendidos. Conhecimento vai além da informação, constitui um saber. Para Turchetii (2011), conhecimento é a habilidade de transformar a informação em material útil para a tomada de decisões.

Adquirir informação tem sido cada vez mais fácil, há muitos meios de comunicação para levar a informação à sociedade, e a tecnologia é um fator muito importante, que associa a informação à comunicação. É a chamada Tecnologia de Informação e Comunicação (TIC). Nos dias de hoje, a maioria das empresas já utilizam a TIC, empresas de grande, médio, e até de pequeno porte buscam estar atualizadas a fim de manter a vantagem competitiva. Para isto, existem sistemas que auxiliam no processo de transformar os dados em conhecimento.

Para a tomada de decisões dentro de uma organização, o processo de transformação dos dados em conhecimento se faz necessário. Com isto, surge a necessidade de ferramentas que auxiliem neste processo, para que seja feito de maneira rápida e fácil. De acordo com Williams (2007, *apud* NOGUEIRA, 2011) os sistemas de *Business Intelligence* (BI) tem alcançado sucesso no mercado empresarial com ferramentas de previsão de custos, relatórios e tomada de decisões.

Segundo Sezões *et. al.* (2006), a implementação de um projeto de *business intelligence* é hoje uma decisão crítica e complexa, que deve ser gerida com o máximo rigor. Desde o momento em que se estudam as necessidades e a justificção de implementação, passando pela avaliação provisional da sua rentabilidade até o processo de implementação no ambiente, são necessárias competências significativas de gestão de projeto, risco, e mudança, para que o resultado seja o esperado.

A partir da década de 90, a Tecnologia da Informação incluía a área de Data Warehouse (INMON, 1997). Em um ambiente de BI, *Data Warehouse* é uma tecnologia que utiliza como matéria-prima as bases de dados legadas, criando repositórios de informações agregadas, visando atender os níveis decisórios das organizações. (NOGUEIRA, 2011).

Em uma instituição de ensino superior onde existem vários sistemas integrados, trabalhar com BI e suas tecnologias facilita o processo da tomada de decisões. Suas tecnologias permitem transformar dados em conhecimento, como também fazer uma análise das informações obtidas para que as decisões tomadas a partir delas não sejam errôneas. Porém, existem organizações que possuem ambiente BI, mas ainda não utilizam de suas tecnologias. Qual a vantagem que as empresas que já as utilizam têm sobre estas?

Os bancos de dados dos sistemas de uma universidade armazenam uma grande quantidade de dados, isto faz com que geralmente se perca muito tempo em consultas às bases. Neste caso, dispor das tecnologias de BI é uma grande vantagem, pois, além de gerar conhecimento, o tempo de resposta às buscas se torna mais curto.

A Universidade Federal da Paraíba, por exemplo, contém uma vasta quantidade de dados dos sistemas que a compõe. Contudo, estes dados ainda não geram conhecimento. Um dos sistemas que a UFPB possui é o chamado Sistema de Controle Acadêmico, em sua base estão armazenados todos os dados referentes ao meio acadêmico da instituição.

Imagina-se que um órgão da UFPB precise fazer uma análise sobre a evasão dos alunos ocorrente em um determinado período de tempo. Utilizar de uma tecnologia que gere conhecimento através dos dados, apresentando resultados analíticos através de relatórios e estatísticas, e de fácil compreensão, seria uma vantagem, pois auxiliaria no processo de tomada de decisão.

Neste contexto, surge uma proposta de aplicação da tecnologia *Data Warehouse* na Superintendência de Tecnologia da Informação (STI) para auxiliar nos processos decisórios correspondentes a Universidade Federal da Paraíba – UFPB.

1.2 OBJETIVOS

1.2.1 Objetivo

Desenvolver uma proposta de aplicação utilizando *Data Warehouse* no ambiente de BI da Superintendência de Tecnologia da Informação da Universidade Federal da Paraíba-UFPB.

1.2.2 Objetivos Específicos

- Extrair, transformar e carregar dados da base de dados do Sistema de Controle Acadêmico da UFPB, para criação e implementação de um *Data Warehouse*.
- Gerar dados estatísticos e relatórios, com o auxílio de uma ferramenta, para analisar dados sobre os alunos do curso de Sistemas de Informação do Campus IV da Universidade Federal da Paraíba, extraindo algumas informações relevantes que auxiliem tomadas de decisão.

1.3 ESTRUTURA DO TRABALHO

Além do capítulo introdutório, este trabalho está composto dos seguintes capítulos:

- Capítulo 2: Este capítulo envolve toda a fundamentação teórica necessária para este trabalho. Nele são apresentados conceitos e práticas importantes de *Business Intelligence*, e das tecnologias *Data Warehouse* e *On-line Analytical Processing*.
- Capítulo 3: Este capítulo apresenta uma proposta de aplicação de *Data Warehouse* no ambiente de BI da Superintendência de Tecnologia da Informação da UFPB, assim como os resultados analíticos obtidos com a implementação do *Data Warehouse*. A ferramenta utilizada para o desenvolvimento desta prática será a *QlikView*.
- Capítulo 4: Neste capítulo são apresentadas as considerações finais do trabalho, com uma proposta de ideias que possam gerar novos trabalhos de pesquisa nesta área.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo serão abordadas as práticas de *Business Intelligence*, a tecnologia *Data Warehouse*, e OLAP (*On-line Analytical Processing*).

2.1 BUSINESS INTELLIGENCE

Business Intelligence (BI), vem sendo cada vez mais utilizado em empresas de pequeno, médio, e grande porte. Nesta seção, serão abordados os conceitos, objetivo, ambiente, processo e tecnologias de BI, para que seja entendido o porquê de sua importância para as empresas.

2.1.1 Contextualização

Segundo Sezões *et al.* (2006), em meados da década de 60, com o surgimento dos primeiros computadores, deu-se início as tarefas de automatização e armazenamento de dados. Por conseguinte, surgiram diversos problemas pela falta de infraestruturas de conexão para troca de dados, assim como sistemas incompatíveis. Com isso, as tarefas levavam muito tempo para serem executadas e eram pouco eficazes. A evolução tecnológica ao longo dos anos fez com que a automatização de processos dentro das organizações se tornasse cada vez mais necessária.

O termo *Business Intelligence* (em Português, Inteligência de Negócios) que surgiu na década de 80 no *Gartner Group*¹, refere-se a um conjunto de aplicações onde através delas se tem um acesso rápido, partilhado e interativo de informações originadas de diversas fontes, para auxiliar no apoio à tomada de decisões dentro das organizações.

“Os sistemas de business intelligence são, na atualidade, os catalisadores da mudança, permitindo concretizar na prática estas novas abordagens de boa gestão e de bom governo das empresas.” (SEZÕES *et al.*, 2006).

Uma vez aplicada dentro de uma organização, BI acrescenta valor em diversos setores. Através dela, é possível fazer análises detalhadas do nível operacional ao estratégico, sendo de muita valia – por vezes essencial – para o processo de tomada de decisões.

Gartner Group é uma empresa de consultoria fundada em 1979 por Gideon Gartner. A *Gartner* desenvolve tecnologias relacionadas a introspecção necessária para seus clientes tomarem suas decisões todos os dias. (ALVES, 2010).¹

De um modo geral, pode-se dizer que BI tem como objetivo transformar informações extraídas de diversas fontes em conhecimento útil. Para isto, ela abriga uma série de tecnologias que são de fundamental importância. As principais são: *Data Warehouse*, *Data Mart* e *Online Analytical Processing (OLAP)*, que serão vistas ao longo do capítulo.

2.1.2 Ambiente de BI

O processo de BI consiste primeiramente em uma coleta de dados originados de diversas fontes (internas e externas) que passam por um processo de ETL que consiste na extração, transformação, e carga de dados. Em seguida, estes dados são armazenados em um repositório chamado *Data Warehouse (DW)*, que consiste em um grande banco de dados histórico. Dentro de um DW encontram-se os *Data Marts (DM)*, pequenos repositórios que são capazes de armazenar dados específicos, pode-se dizer que os DM são partes de um grande repositório, o Data Warehouse. Uma vez coletados, extraídos, transformados, e carregados para um grande repositório, os dados são resultados em conhecimento útil, e através de relatórios ou sistemas OLAP (*On-line Analytical Processing*) que são aplicações que permitem facilitar a análise das informações originadas do DW, auxiliam na tomada de decisões dentro de uma organização. A Figura 01 ilustra o processo descrito acima.

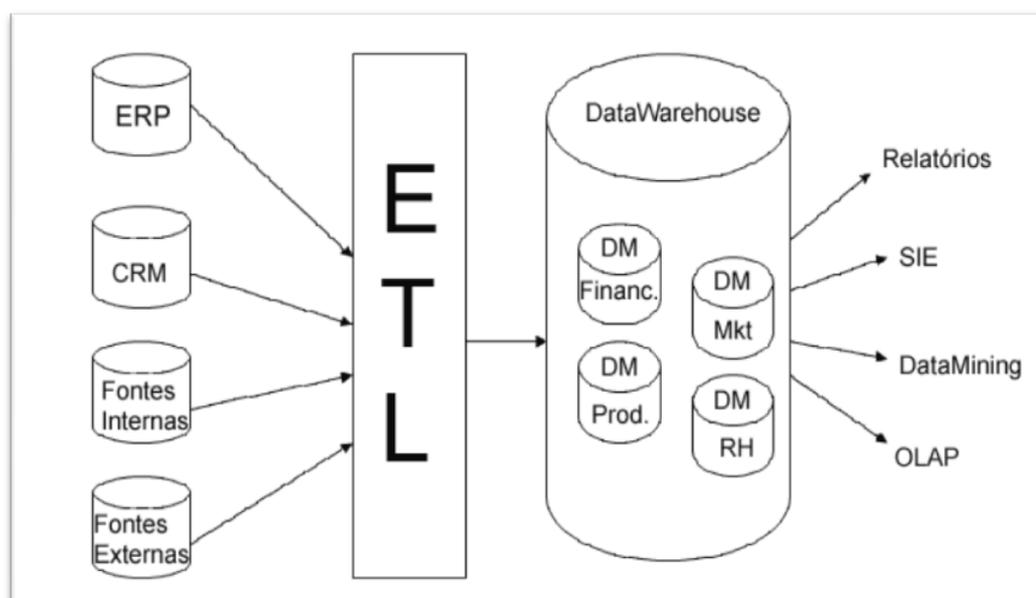


FIGURA 01: Arquitetura de BI
FONTE: Kaupa (2008)

2.1.3 Extração, Transformação, e Carga (ETL)

O processo de ETL (*Extract, Transform, e Load*), como o próprio nome já diz, é o processo destinado à extração, transformação e carga dos dados. Nele, os dados são extraídos de fontes externas ou internas para serem carregados em uma ou mais bases de dados Data Warehouse. É um processo minucioso, onde os dados precisam ser atentamente analisados e se preciso transformados passando por meio de um processo de padronização, para que ao chegar ao DW os mesmos estejam uniformes e com uma forma de representação única, evitando assim que as informações geradas posteriormente não sejam errôneas.

“A etapa de ETL é uma das mais críticas de um *Data Warehouse*, pois envolve a fase de movimentação dos dados.” (PRIMAK, 2008 p. 63 *apud* ANTONELLI, 2009).

Abaixo segue a descrição das três etapas deste processo:

Extract (extração) é a etapa responsável pela busca de dados de fontes heterogêneas, que podem ser externas ou internas, de um ou mais banco de dados. Após coletados, os dados são armazenados na *Staging area*, área esta em que ocorre a validação e limpeza dos dados antes de serem armazenados no DW.

Transform (transformação) é a etapa responsável por transformar os dados extraídos, com o intuito de padronizá-los em um só formato antes de serem carregados no DW. Essa transformação se dá em relação ao tamanho e tipo, correções ortográficas, verificação de duplicidade de informações, problemas de acentuação, dentre outros.

Load (carga) última etapa do processo, é a responsável pela carga dos dados no DW. Nela os dados que foram extraídos são tratados a partir das análises detalhadas, para assim serem carregados no DW. Esta carga pode ser executada semanalmente, diariamente, ou até várias vezes por dia, isto varia de acordo com a necessidade da organização.

Segundo Cielo (2008), existem ferramentas como *Data Stage* da IBM, *ETI* da ETI corporation, Pentaho (*Open Source*) e *Oracle Warehouse Builder* da Oracle que auxiliam no processo de ETL.

2.2 DATA WAREHOUSE

Como visto anteriormente, após o processo de ETL os dados são carregados no *Data Warehouse* (DW), para que através dele, sejam geradas as informações úteis para a tomada de decisões dentro de uma organização.

Data Warehouse (em Português, Armazém de Dados) é uma estrutura de banco de dados onde são armazenadas informações importantes para o processo de BI. Nele os dados

analíticos são armazenados e resultados em consultas e gráficos que auxiliarão na tomada de decisões dentro de uma organização, fornecendo dados integrados e históricos. Para os especialistas da área da informática o *Data Warehouse* é como uma “peça” importante para a execução prática de um projeto de BI (PRIMAK, 2008 *apud* MEDEIROS, 2011).

De acordo com Kimball et. al. (1998), *Data Warehouse* é uma fonte de dados consultáveis da organização, formada pela união de todos os *data marts* correspondentes.

Segundo Inmon (1997, *apud* DOMENICO, 2001), *Data Warehouse* é uma coleção de dados orientados por assuntos, integrados, variáveis com o tempo e não voláteis, para dar suporte ao processo gerencial de tomada de decisão. Com base nesta definição, a seguir serão abordadas as principais características do DW.

2.2.1 Características

Orientado por assunto: Um *Data Warehouse* armazena informações de acordo com o interesse das pessoas que irão fazer uso delas, e isto varia em cada organização. Por exemplo, em uma instituição de ensino superior, um chefe de departamento quer saber a situação dos alunos de determinados cursos, para isto, devem ser definidas quais as informações importantes para a análise.

Integrado: Em sistemas transacionais, os dados podem estar representados de diversas formas. Isto geralmente ocorre com os gêneros “masculino” e “feminino”, pois podem ser encontrados nas bases de formas diferentes, como por exemplo: ‘M’, ‘F’; ‘1’, ‘2’; ‘H’, ‘M’. Com a integração dos dados no DW, os mesmos são padronizados para que sejam representados de uma única forma.

Não volátil: Um *Data Warehouse* é não volátil, pois, ao contrário dos sistemas transacionais, nele os dados não podem ser alterados ou atualizados depois de registrados. Ou seja, uma vez carregados no DW, os dados se tornam acessíveis apenas para leitura, exceto quando ocorrem cargas ou inclusão de novos registros.

Variante no tempo: Em um DW, dados não são perdidos após uma atualização, a cada ocorrência de uma mudança, um novo registro é criado para marcar esta mudança e os antigos continuam armazenados. A determinação do tempo pode assumir muitas formas, desde uma marcação de tempo em todos os registros, a uma marcação para uma base de dados inteira (INMON, 2007). Estas marcações podem ser anuais, mensais, semanais, ou até diárias, tendo maior grau de complexidade nas duas últimas devido às variações encontradas no número de dias em um mês ou em um ano, ou ainda no início das semanas dentro de um mês.

Granularidade: Segundo Inmon (2007), um aspecto muito importante de projeto de *Data Warehouse* é a questão da granularidade. A mesma refere-se ao nível de detalhes ou de resumos contidos nos dados existentes em um DW. Seu nível varia de acordo com o nível de detalhes (sumarização) dos dados, ou seja, quanto mais detalhado, menor o nível de granularidade, quanto menos detalhado, maior o nível de granularidade. O nível de granularidade afeta diretamente o volume de dados armazenados no DW e ao mesmo tempo, o tipo de consulta que pode ser respondida.

De acordo com Teixeira (2013), para um *Data Warehouse* ser útil, deve ser capaz de responder a consultas avançadas de modo rápido, sem deixar de exibir detalhes relevantes à resposta. Para isso, deve possuir uma arquitetura que nos permita coletar, manipular e apresentar os dados de maneira eficiente e rápida.

2.2.2 Arquitetura

A definição de uma arquitetura básica de *Data Warehouse* facilita o entendimento do processo e de todas as tecnologias envolvidas em seu desenvolvimento. Na Figura 02, é mostrada uma visão geral desta arquitetura.

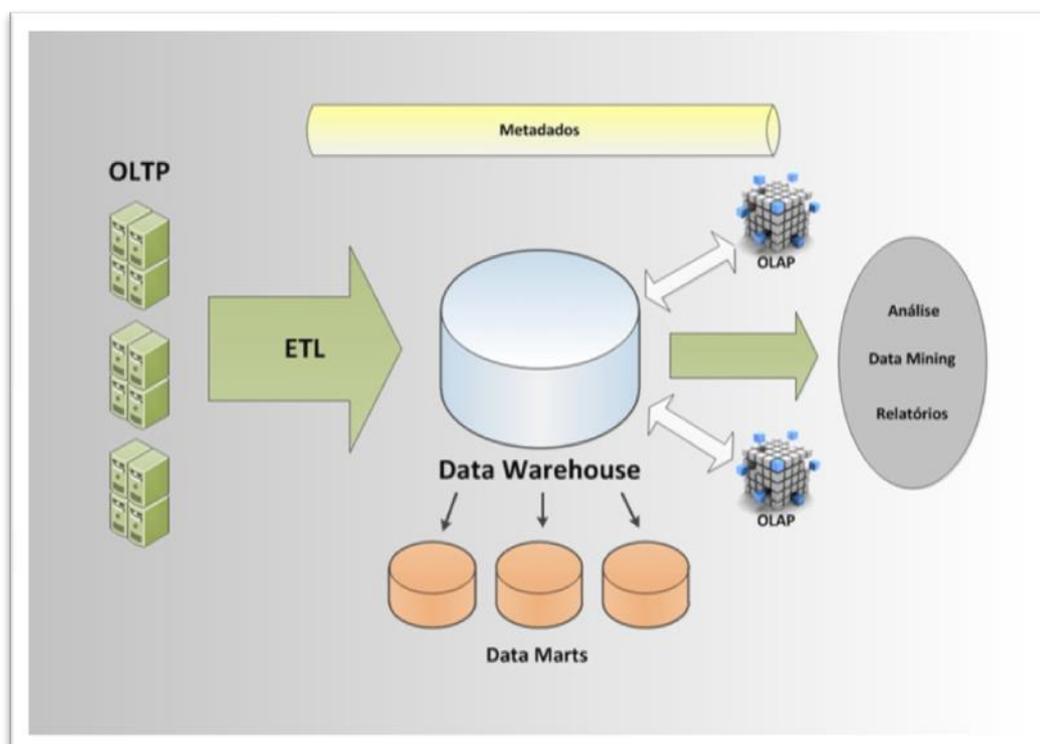


FIGURA 02: Arquitetura básica de um DW
FONTE: Cruz (2011)

Como visto na figura acima, a arquitetura básica de DW é composta basicamente pelos processos OLTP (*Online Transaction Processing*), ETL (*extract, transform, load*), *Data Marts*, Metadados, e *Online Analytical Processing*, conhecida como OLAP. A seguir, um breve resumo de como esses componentes interagem com o *Data Warehouse*.

O processamento de transações em tempo real (OLTP) é responsável por trazer dados de fontes externas para serem carregados no DW. Antes de serem carregados, estes dados são extraídos das fontes OLTP e transformados para terem seu formato padronizado, com isto se dá o processo de ETL. Para Cruz (2011), *Data Marts* (DM) é um subconjunto de dados de um DW onde geralmente são armazenados dados referentes a um assunto em especial ou diferentes níveis de sumarização, que focalizam uma ou mais áreas específicas. Por armazenar informações mais específicas de uma determinada área ou assunto, as buscas realizadas em um DM podem ser mais detalhadas.

Após os dados estruturados e carregados no DW, são necessárias ferramentas especializadas em leituras para facilitar a interpretação dos mesmos. Estas ferramentas são as aplicações de processamento analítico em tempo real, conhecidas como OLAP. Ainda neste capítulo, a mesma será abordada com mais detalhes.

Uma arquitetura global de DW pode ser estruturada fisicamente de forma centralizada ou distribuída. Centralizada quando a organização existe em um único local e os dados se encontram em uma única instalação física. Distribuída quando a organização tem vários locais físicos, e seus dados podem ser encontrados em múltiplas instalações.

Segundo Cruz (2009), atualmente existem várias propostas de arquiteturas de DW. A seguir serão abordadas as seguintes arquiteturas: *top-down* e *bottom-up*.

2.2.2.1 Arquitetura *Top-Down*

Proposta por Inmon (1997), esta arquitetura permite que um projetista construa um DW a partir das bases de informações, para depois os dados serem carregados nos *Data Marts*. Este tipo de arquitetura pode facilitar na manutenção, partindo da premissa de que os *Data Marts* são sub-conjuntos de um DW. Em contrapartida, a mesma tem uma implementação mais demorada. O modelo *top-down* apresenta alto grau de dificuldade no planejamento e implementação de um modelo único, não diferenciável e mestre a toda

organização (KIMBALL *et al.* 1998, *apud* DOMENICO, 2001). Para melhor entendimento, a estrutura desta arquitetura será mostrada na Figura 03.

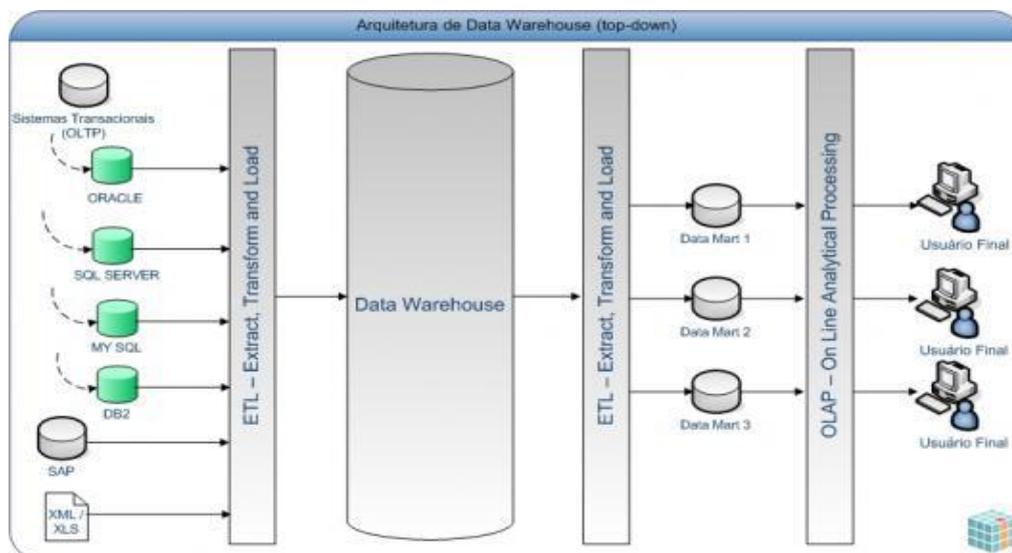


FIGURA 03: Arquitetura *Top-Down*
FONTE: Bonomo (2009)

2.2.2.2 Arquitetura *Bottom-up*

Proposta por Kimball (1998), esta arquitetura permite que o DW seja construído de forma incremental. Isto é possível, pois, neste modelo, os *Data Marts* são construídos independentes do DW. Diferente do modelo *top-down*, neste tipo de arquitetura o desenvolvimento do DW é mais rápido. Outro ponto positivo é que os assuntos de maior interesse dos usuários finais podem ser desenvolvidos inicialmente. Por sua implementação ser feita em paralelo, se faz necessário a utilização de esforços e recursos de várias equipes, isto pode ser considerado uma desvantagem. A estrutura desta arquitetura é mostrada na Figura 04, para melhor entendimento.

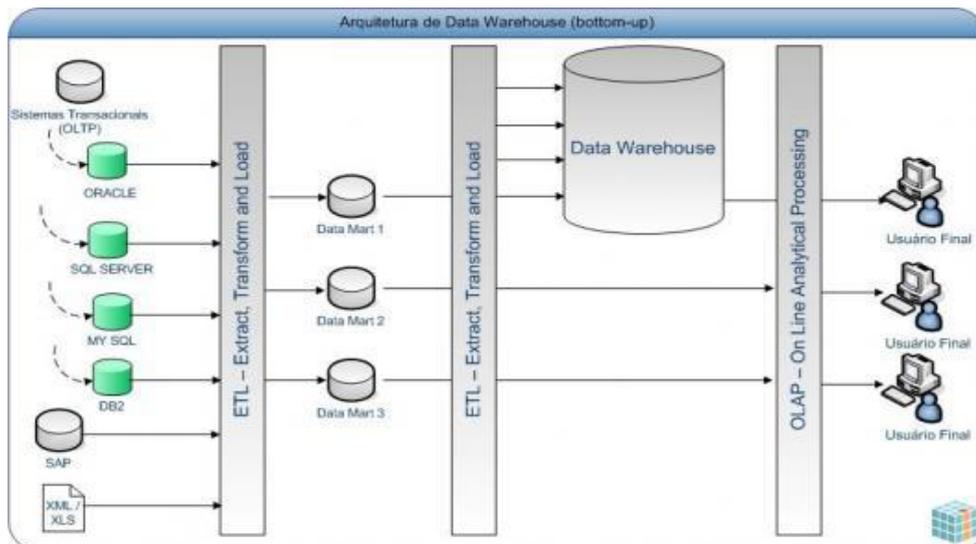


FIGURA 04: Arquitetura *Bottom-up*
FONTE: Bonomo (2009)

2.2.3 Modelo de Dados

A modelagem de dados é considerada um passo essencial para a estrutura de um sistema de informação em um *data warehouse*. Para que um sistema seja bem modelado, é importante que as entidades lógicas sejam coerentes com os processos de negócio da organização, assim como a criação de suas dependências. Em Banco de Dados existem dois modelos de dados bastante conhecidos: *modelo relacional* e *modelo dimensional*, sendo este último o mais utilizado em um processo de DW. A seguir, serão abordados estes dois modelos, dando ênfase ao modelo *dimensional*.

2.2.3.1 Modelo Relacional

De acordo com Rohden (2009), o modelo relacional se baseia no princípio de que todos os *dados* estão dispostos em *tabelas* e toda sua definição é fundamentada na lógica de predicados e na teoria dos conjuntos. Estas tabelas podem se relacionar entre si, isto se dá através dos *relacionamentos*. Tabelas podem ser chamadas de *entidades*, e os dados de *atributos*. Abaixo seguem as definições destes três princípios do modelo relacional.

- *Entidades:* Em um sistema, entidades podem representar objetos ou coisas existentes no mundo real. Como por exemplo: um lugar, uma pessoa, entre outros. Para um bom entendimento na modelagem, as entidades precisam ser nomeadas de forma clara e objetiva.
- *Atributos:* Representam as características de uma determinada coisa ou objeto. Por exemplo, em um sistema acadêmico, na entidade Alunos, os atributos

podem ser: CPF, nome, data de nascimento, entre outros. Em uma entidade existe pelo menos um atributo com valor único. No exemplo de Alunos, o atributo CPF é considerado de valor único, pois, o mesmo varia para cada aluno.

- *Relacionamentos*: São as ligações entre as entidades, geralmente representados por verbos. Nas entidades Alunos e Disciplina, um relacionamento entre elas pode ser representado por *possui*, isto quer dizer que uma disciplina possui aluno.

Segundo Kimball (1998, *apud* FELTES, 2010), o modelo relacional divide os dados em várias entidades distintas, sendo criada uma tabela no banco de dados para cada entidade. Isso gera um imenso número de conexões possíveis entre as tabelas através de junções internas entre os elementos de dados.

Embora o modelo relacional seja bastante usado em sistemas transacionais, em um ambiente de DW onde os dados gerados fazem parte do processo de decisão, é necessário usar um modelo que demonstre aos usuários finais de uma forma simples e de fácil entendimento, a estrutura dos dados que foi construída. Para isto, é utilizado o *modelo dimensional*.

2.2.3.2 Modelo Dimensional

“Modelagem dimensional é uma técnica de design de banco de dados projetada para suportar consultas de usuários finais em um Data Warehouse.” (KIMBALL et. al., 1998).

Segundo Moreira (2006), o modelo dimensional permite visualizar dados abstratos de forma simples e relacionar informações de diferentes setores da empresa de forma muito eficaz, além disso, permite que os *softwares* naveguem pelos bancos de dados com eficiência. Para Novais (2010), um dos objetivos do modelo dimensional é permitir ao usuário realizar consultas na base de dados sem depender da equipe de tecnologia.

Para Adamson (2010 *apud* NOGUEIRA, 2011), a técnica de Modelo Dimensional ajuda na análise de negócios de uma maneira simples: ela modela a medida dos processos de negócios.

Neste modelo as informações podem ser obtidas de várias fontes (planilhas, arquivos, entre outros), e as relações destas informações podem ser representadas em forma de um cubo.

Cubo de Informações

No modelo dimensional, um cubo de informações representa as informações e suas relações em cada eixo. Estas informações podem ser fatiadas e aprofundadas em cada dimensão ou eixo para que sejam extraídos mais detalhes sobre os processos internos de uma determinada organização (MOREIRA, 2006). Em um modelo relacional onde as informações são obtidas através de consultas estruturadas, esta extração torna-se muito complicada.

Considerando o exemplo de que um chefe de departamento precisa fazer uma análise dos cursos do departamento em questão. Neste caso, podem ser consideradas como principais dimensões do cubo: cursos, coordenadores. A Figura 05 mostra estas dimensões no cubo, para melhor entendimento.

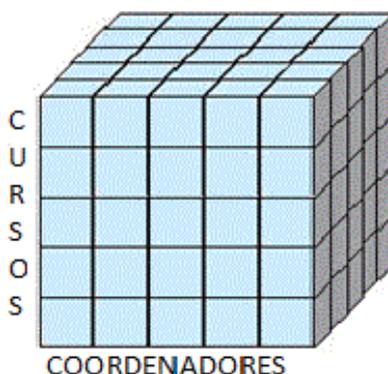


FIGURA 05: Cubo de Informações

FONTE: Adaptado pelo próprio autor segundo Moreira (2006)

A cada eixo do cubo estão as informações contidas nas duas dimensões, ou seja, os eixos representam as intersecções entre as dimensões. Para obter informações ainda mais detalhadas, podem-se acrescentar mais dimensões ao cubo. Um modelo dimensional pode ter quantas dimensões forem necessárias (MOREIRA, 2006).

De acordo com Moreira (2006), outro fator importante para a modelagem dimensional é a velocidade de acesso a uma informação, com modelos simples sem muitas tabelas para relacionar, é muito rápido para extrair as informações necessárias.

2.2.3.2.1 Fatos e Dimensões

Um modelo dimensional consiste em uma tabela central de *fatos* e outras de *dimensões* ligadas diretamente a ela. A seguir, conceitos sobre fatos e dimensões para melhor entendimento.

Fatos

Segundo Novais (2012), na modelagem dimensional, a palavra "fato" representa uma medida dos processos modelados, como quantidades, valores e indicadores. São resultantes de um processo de negócio ou eventos, e geralmente possuem valores numéricos. Uma tabela de fatos registram os fatos que serão analisados, a mesma deve conter atributos referentes a um fato específico, ou seja, não se deve misturar vários assuntos em uma mesma tabela.

Para Kimball (1998, *apud* NOVAIS, 2012) a tabela de fatos é a principal tabela de um modelo dimensional, onde as medições numéricas de interesse da empresa estão armazenadas.

Existem informações armazenadas nas tabelas de fatos que permitem medir o desempenho dos processos de negócio, estas informações são denominadas de *métricas*. Métricas podem ser volumétricas, numéricas, e na maioria das vezes são aditivas, ou seja, podem permitir operações matemáticas como: adição, subtração e médias.

Dimensões

“Dimensões é o núcleo de componentes ou categorias do processo de negócio, ou seja, tudo que se quiser analisar nos relatórios ou visões.” (LIMA, 2010).

Tabelas de dimensões são compostas por atributos que descrevem um determinado negócio, e geralmente são tabelas bem menores que uma tabela de fatos. Para Novais (2012), uma tabela de dimensões possui aspectos pelos quais se pretende observar métricas relativas ao processo modelado.

Segundo Lima (2010), as Dimensões são consideradas objetos de negócio e os Fatos são medidas de negócio. No cenário de uma empresa de vendas, por exemplo, os objetos podem ser os produtos, os clientes, as empresas, etc., e as medidas podem ser as quantidades pedidas, os valores vendidos, etc. Produtos existem, mesmo sem serem vendidos, mas as vendas sempre são de produtos ou serviços.

Para melhor entendimento, a seguir será criado um cenário onde se utiliza uma tabela de fatos e suas respectivas dimensões com base nos conceitos acima.

- Em uma empresa comercial, o gerente quer fazer uma análise das vendas realizadas em um determinado período para saber se obteve ou não o lucro esperado. Para isto, primeiramente deve-se identificar qual o “fato” a ser analisado. Em seguida, descrevê-lo através de suas “dimensões”.

Neste contexto, pode-se identificar o fato como “Vendas”, pois, é o que espera ser analisado. As dimensões deste fato serão tudo aquilo que descreve este processo de negócio. Podem-se identificar como principais dimensões: o tempo (o período em que as vendas foram realizadas), os produtos que foram vendidos, as lojas, e os clientes. A Figura 06 mostra um exemplo de modelo dimensional para este caso, com a tabela central de fatos representada por ‘Fatos Vendas’ e as tabelas dimensionais correspondentes: ‘Dimensão Tempo’, ‘Dimensão Produto’, ‘Dimensão Loja’, e ‘Dimensão Cliente’ que se relacionam diretamente com a tabela de fatos. É importante observar que as chaves primárias (*Primary Keys*) das tabelas de dimensões são identificadas por chaves estrangeiras (*Foreign Keys*) na tabela de fatos. Esta identificação funciona como referência das dimensões na tabela do fato em questão.

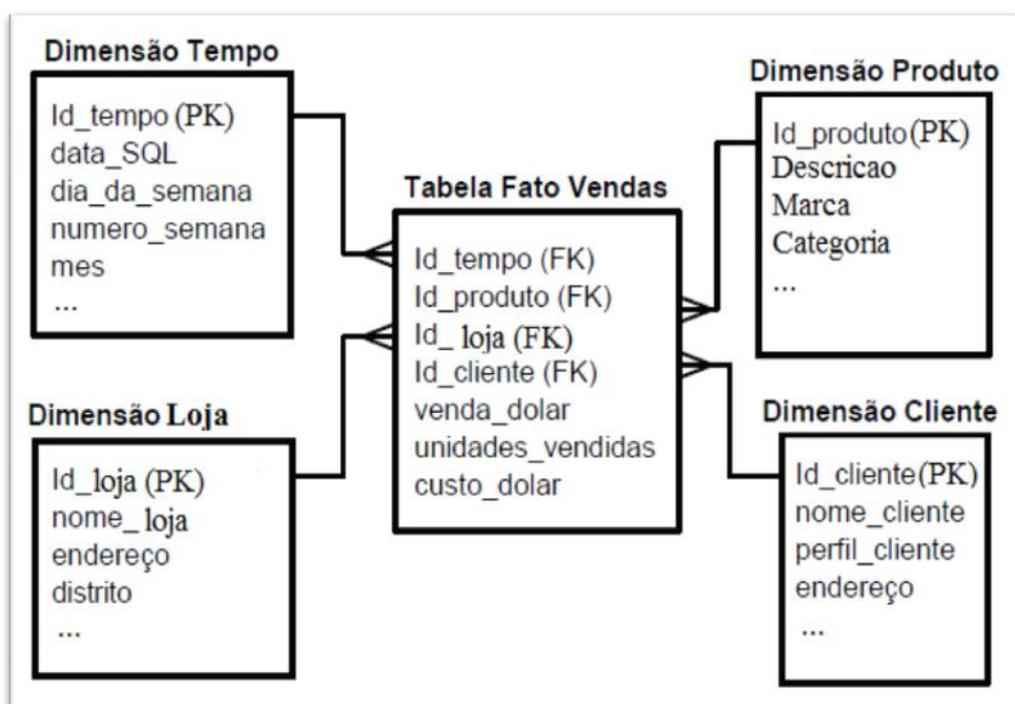


FIGURA 06: Exemplo de Fatos e Dimensões em um processo de Vendas

FONTE: Adaptado pelo próprio autor segundo Wagner (2012)

É importante observar também que, além das referências das dimensões na tabela de Fatos, a mesma também possui suas próprias informações. No exemplo acima os atributos ‘venda_dolar’, ‘unidades_vendidas’, e ‘custo_dolar’ contém valores numéricos que podem ser aditivos, ou seja, calculados por meio de operações matemáticas. As dimensões por sua vez contém atributos que são características próprias de cada uma delas.

A modelagem dimensional possui modelos de dados, dois desses modelos são: *Star Schema* (Modelo Estrela) e *Snow Flake* (Floco de Neve).

2.2.3.2.1.1 *Star Schema* (Modelo Estrela)

Chamado de modelo Estrela por conter uma tabela de Fatos centralizada e as tabelas dimensionais ligadas a ela em um formato assemelhado ao de uma estrela. Nele, as tabelas dimensionais se relacionam diretamente com a tabela de Fatos. Uma característica importante desse modelo é que nele as tabelas de dimensões não são normalizadas para evitar *joins* entre as tabelas, ou seja, as tabelas de dimensões contêm descrições que se repetem a cada registro, isso faz com que o tamanho das tabelas aumente em cada registro. A Figura 07 mostra o modelo Estrela também de um processo de Vendas como na figura anterior, onde a tabela central é representada por ‘Fatos_Vendas’ diretamente relacionada com as tabelas dimensionais que são representadas por: ‘Dimensão Marketing’, ‘Dimensão Produto’, ‘Dimensão Tempo’, ‘Dimensão Cliente’, ‘Dimensão Loja’ e ‘Dimensão Promoção’.

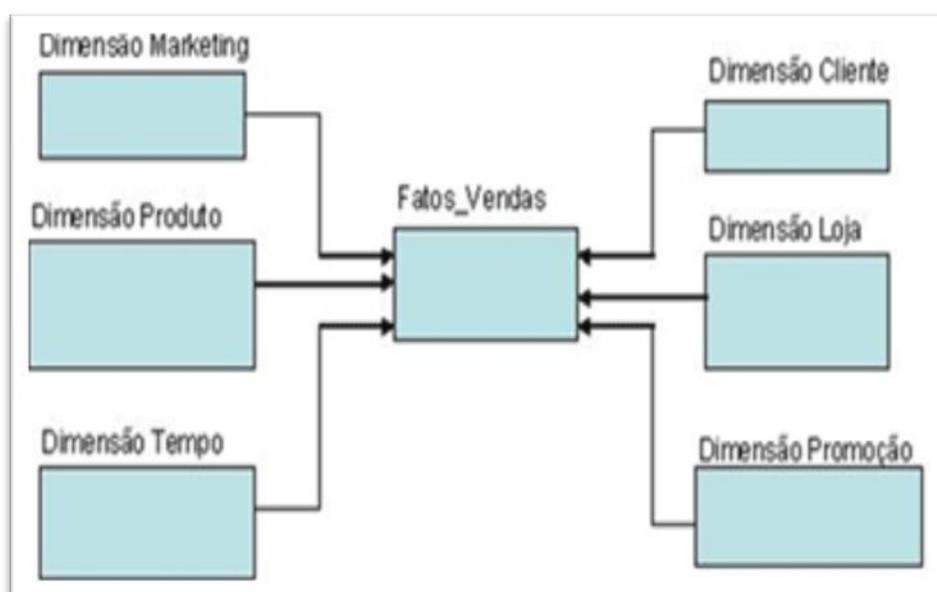


FIGURA 07: *Star Schema* ou Modelo Estrela

FONTE: Adaptado pelo próprio autor segundo Moreira (2006)

2.2.3.2.1.2 *Snow Flake* (Floco de Neve)

Neste modelo as tabelas de dimensões também se relacionam com a tabela de Fatos, porém, além disto, também se relacionam entre elas mesmas. Isto ocorre para normalizar as tabelas dimensionais visando diminuir o espaço ocupado pelas mesmas, o que não acontece no modelo Estrela, onde as tabelas de dimensões não são normalizadas. Outra característica deste modelo, é que podem existir tabelas dimensionais auxiliares para normalizar as tabelas de dimensões principais. A Figura 08 mostra um exemplo de um modelo Floco de Neve que representa o mesmo processo de negócio da Figura 07. A diferença deste modelo é que além

das tabelas dimensionais estarem diretamente relacionadas com a tabela ‘Fatos Vendas’, as tabelas ‘Dimensão Produto’ e ‘Dimensão Tempo’ possuem tabelas auxiliares relacionadas a elas.

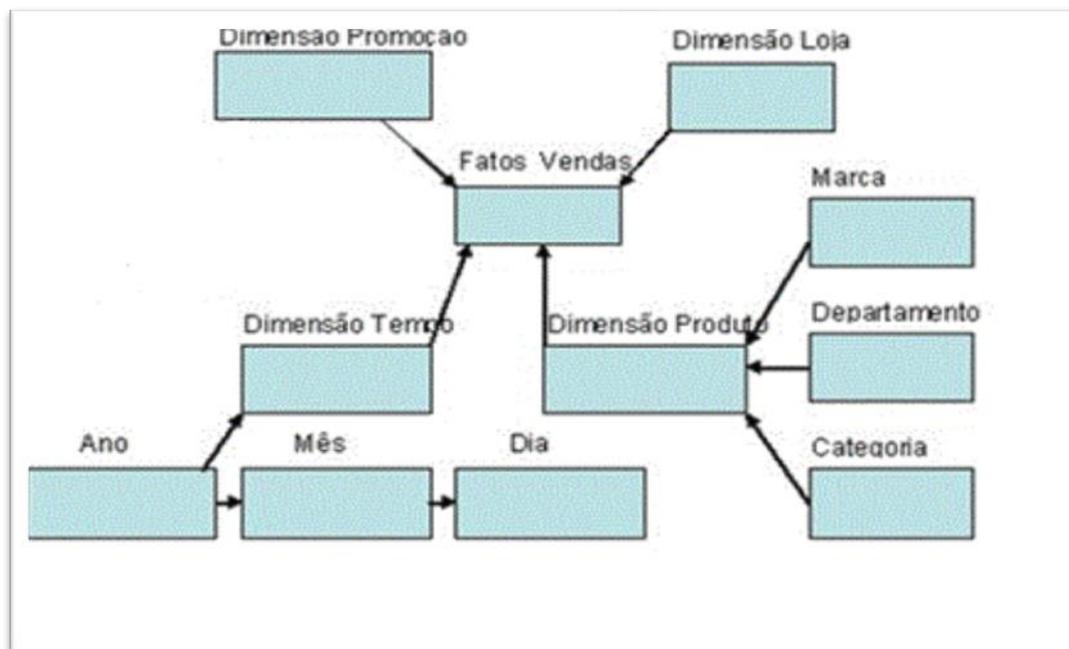


FIGURA 08: Modelo *Snow Flake* ou Floco de Neve

FONTE: Adaptado pelo próprio autor segundo Moreira (2006)

Segundo Moreira (2006), é preferível utilizar o modelo Estrela ao modelo Floco de Neve, pois, o primeiro fornece acesso mais rápido aos dados, uma navegação mais fácil, e cria tabelas auxiliares somente para dimensões específicas.

2.2.4 *On-line Analytical Processing (OLAP)*

Segundo Sezões *et. al.* (2006), OLAP é um conceito que se refere a aplicações informáticas que permitem efetuar, de forma rápida e partilhada, a análise de informação multidimensional, originária de diversas fontes de dados. As tecnologias OLAP fazem com que os dados de uma base relacional sejam reestruturados em uma concepção multidimensional.

Uma vantagem no OLAP é a rapidez com que suas análises são feitas, por exemplo, questões que demorariam alguns dias para serem respondidas nas bases de dados tradicionais podem ser solucionadas em apenas alguns segundos no processamento OLAP.

De acordo com Caldas et. al. (2006), as aplicações baseadas em OLAP também podem ser chamadas de aplicações sobre o negócio. Estas aplicações analisam as informações obtidas através das aplicações operacionais, ajudando executivos e altos gerentes a interpretar mudanças na realidade de negócio, e, assim, mudar as regras de negócio da empresa.

2.2.4.1 Arquiteturas OLAP

Aplicações OLAP podem ser estruturadas de diversas formas. Algumas delas são:

ROLAP (*relational OLAP*): Caracterizado por manter os dados nas tabelas relacionais originais, ao mesmo tempo em que gera outras tabelas onde não são armazenados os valores agregados. Estes dados agregados são somas com um baixo nível de detalhe derivadas dos dados. (SEZÕES *et. al.*, 2006). Um exemplo é a agregação de valores que representam unidades de tempo como: dia, semana, mês, etc. Esta agregação de dados é considerada uma vantagem no ROLAP, pois, permite o aumento do desempenho, em contrapartida, por sustentar um modelo relacional, esta solução se torna mais lenta que as demais.

MOLAP (*multidimensional OLAP*): Diferente do ROLAP, este tipo armazena as informações com uma estrutura de dados multidimensional, e não gera outras tabelas para armazenar valores agregados. Tudo é calculado automaticamente como parte da própria modelagem dimensional. Uma vantagem no MOLAP é a rapidez das respostas às questões feitas pelos usuários, em contrapartida, esta solução ocupa um grande espaço de memória.

HOLAP (*hybrid OLAP*): Este modelo envolve as duas arquiteturas anteriores. Utiliza de tabelas relacionais para guardar os dados, e armazena as agregações como uma estrutura multidimensional. Para Sezões et. al. (2006), o HOLAP assume-se como uma opção de compromisso, quer em termos de rapidez de resposta, quer em termos de dimensão.

DOLAP (*desktop OLAP*): Segundo Feltes (2010), o objetivo deste modelo é facilitar a *performance* de alguma análises, diminuindo o tráfego de informações cliente-servidor. Para isto, utiliza-se uma abordagem onde as estruturas dimensionais e relacionais são transferidas para as estações clientes. Esta abordagem é mais aplicada em bancos de dados individuais utilizados para análises mais detalhadas.

2.2.4.2. Operações OLAP

Para Singh (2001, *apud* FELTES, 2010), as ferramentas OLAP fornecem subsídios para a análise complexa do negócio, possibilitando aos usuários analisar e navegar pelos dados para identificar tendências, exceções e obter alguns detalhes para entender melhor os

resultados do seu negócio. Para isto, estas ferramentas contam com algumas operações OLAP. São elas:

Drill Down: Ocorre quando há movimentação dos dados do nível mais alto para os dados mais detalhados de uma dimensão, diminuindo assim o nível de granularidade. A Figura 09 mostra um exemplo da navegação dos dados em um cubo de informações com *Drill Down*. As siglas apresentadas na vertical do primeiro cubo representam os Campi de uma universidade, e à direita estão representados os anos, isto é, o cubo contém informações de cada Campus nos respectivos anos de '2011', '2012', e '2013'. No segundo as informações encontram-se mais detalhadas, cada fatia do cubo representam meses do ano de '2013'. Pode-se observar que a navegação se dá do primeiro ao segundo cubo, ou seja, do nível mais alto ao mais detalhado dentro das dimensões.

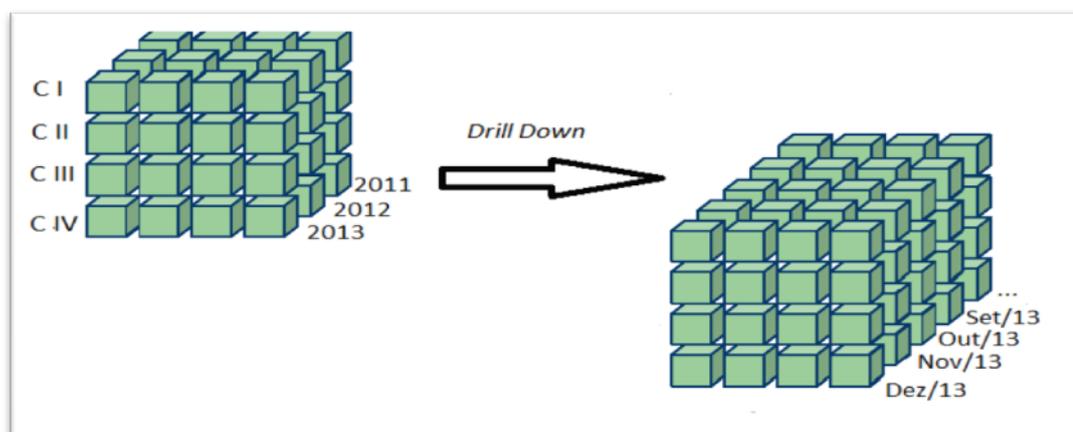


FIGURA 09: Exemplo de *Drill Down* no cubo de informações
FONTE: Adaptado pelo próprio autor segundo Timóteo (2012)

Drill Up/ Roll Up: O contrário do *Drill Down*. Ocorre quando há movimentação dos dados mais detalhados ao nível mais alto em uma dimensão, aumentando assim o nível de granularidade. A Figura 10 mostra o mesmo exemplo de cubo de *Drill Down*, a grande diferença é que neste pode-se observar que a navegação se dá do segundo ao primeiro cubo, ou seja, vai do nível mais detalhado ao mais alto dentro das dimensões.

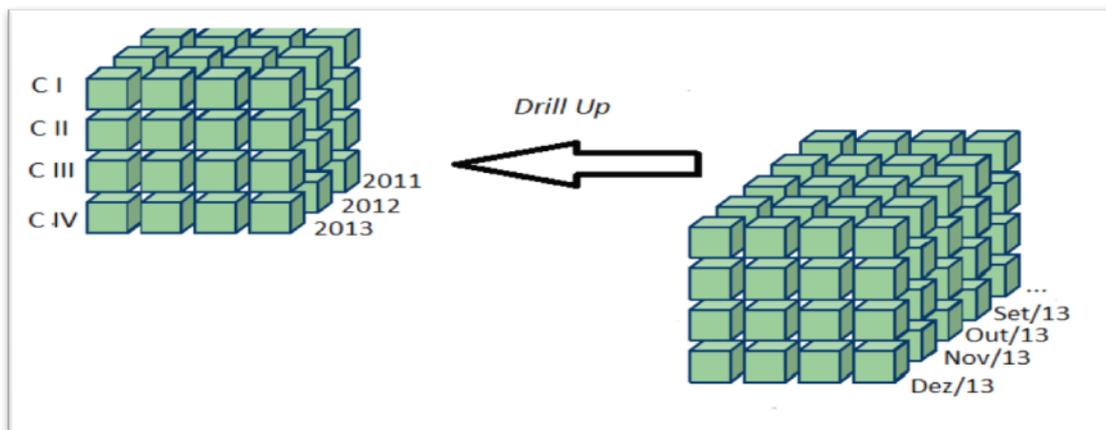


FIGURA 10: Exemplo de *Drill Up/Roll Up* no cubo de informações
FONTE: Adaptado pelo próprio autor segundo Timóteo (2012)

Drill Across: Permite a navegação dentro de uma tabela de dimensão passando de um nível mais alto ao mais detalhado sem passar pelos níveis intermediários. A Figura 11 mostra uma pirâmide que representa o nível mais ao alto ao mais detalhado dentro de uma dimensão. Nela a busca das informações é feita de “Ano” a “Dia” sem passar por “Mês” e “Semana”.

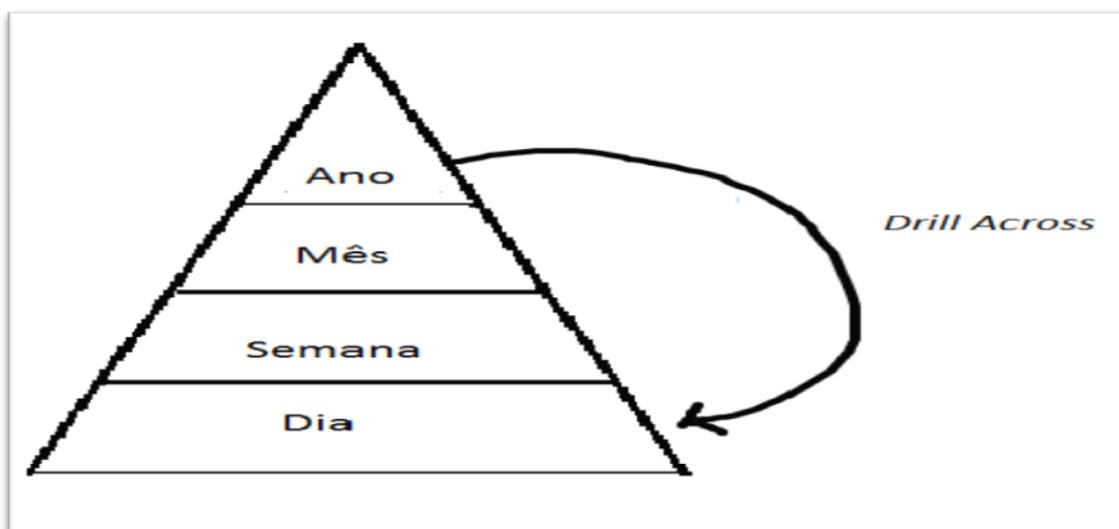


FIGURA 11: Exemplo de *Drill Across*
FONTE: Adaptado pelo próprio autor segundo Medeiros (2011)

Drill Through: Ocorre quando a navegação passa de uma informação contida em uma dimensão para outra. A Figura 12 mostra um exemplo de *Drill Through*, onde um usuário busca por um nome na dimensão ALUNO, e depois passa a buscar um código na dimensão DISCIPLINA.

| | Nome | Nota |
|-----------------------|----------------------|------|
| A L U N O | Maria Eduarda | 9,5 |
| | <u>Bruna Leticia</u> | 8,0 |
| | Lucas Andrade | 6,5 |

Drill Through

| | Código | Carga Horária |
|------------|-----------------|---------------|
| DISCIPLINA | 08333120 | 060 |
| | <u>08333203</u> | 060 |
| | 08333678 | 060 |

FIGURA 12: Exemplo de *Drill Through*

FONTE: Adaptado pelo próprio autor segundo Medeiros (2011)

Slice and Dice: Operação que fatia o cubo de informações com o objetivo de facilitar a compreensão dos usuários na navegação dos dados. De acordo com Oliveira; Pereira (2010), *Slice* é a operação que corta o cubo, mas mantém a mesma perspectiva de visualização dos dados. *Dice* é a mudança de perspectiva da visão. É a extração de um subcubo ou a intercessão de vários *slices*. As Figuras abaixo mostram exemplos de *Slice* e *Dice* retomando os exemplos dos cubos de *Drill Down* e *Drill Up* mostrados anteriormente. Na Figura 13 pode-se observar que a “fatia” que representa o ano de 2012 é cortada do cubo para que o usuário possa navegar apenas pelas informações contidas na mesma. Já na Figura 14 o exemplo mostra uma mudança de perspectiva de visão do cubo. Nota-se que ao fazer vários *slices*, isto é, quando as fatias que representam “C III”, “2012”, e “B” (neste exemplo as letras de ‘A a D’ representam setores dentro dos determinados Campi) são extraídas para que o usuário possa navegar pelas informações nelas contidas, um pequeno cubo é apresentado. O pequeno cubo é considerado um subcubo ou uma intersecção entre estas dimensões.

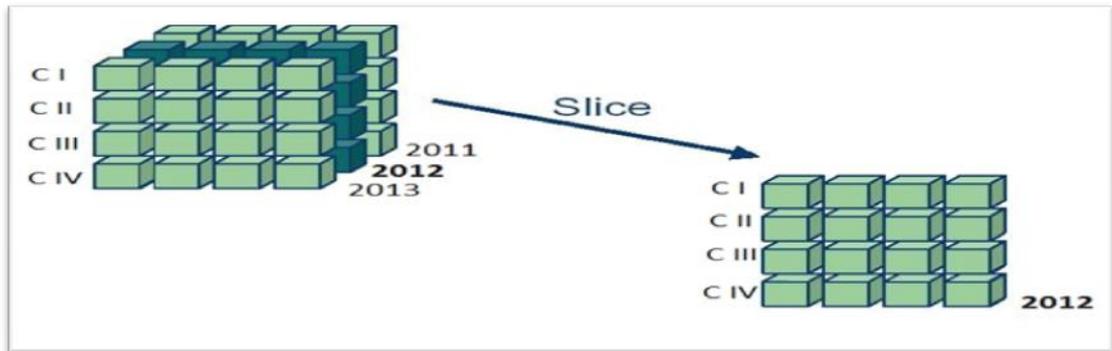


FIGURA 13: Exemplo de *Slice* no cubo de Informações
FONTE: Adaptado pelo próprio autor segundo Timóteo (2012)

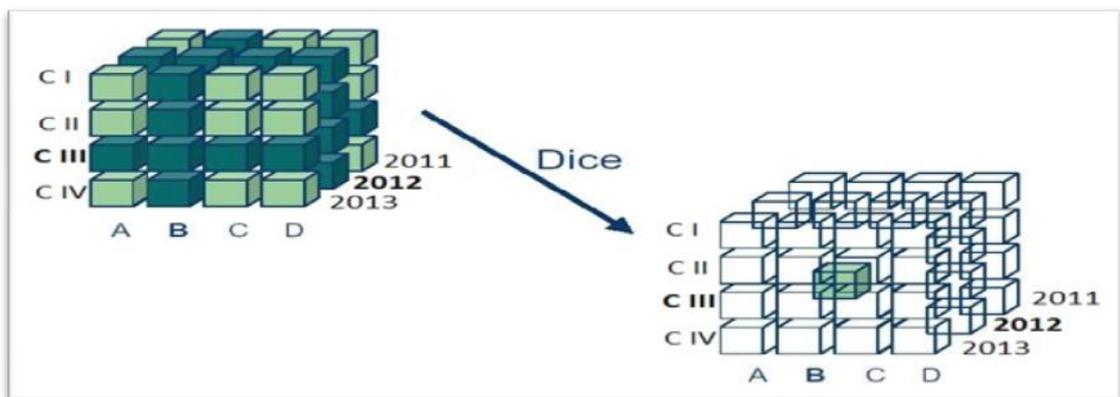


FIGURA 14: Exemplo de *Dice* no cubo de Informações
FONTE: Adaptado pelo próprio autor segundo Timóteo (2012)

Pivoting: Operação que permite a mudança de posições das linhas e colunas dentro de uma tabela dimensional, a fim de se obter uma nova visão das informações. A Figura 15 mostra um exemplo de tabelas onde as informações nelas contidas mudam de posições. Nota-se que as linhas da primeira tabela viram colunas na segunda, e as colunas viram linhas. Isto permite ao usuário uma nova visão das informações.

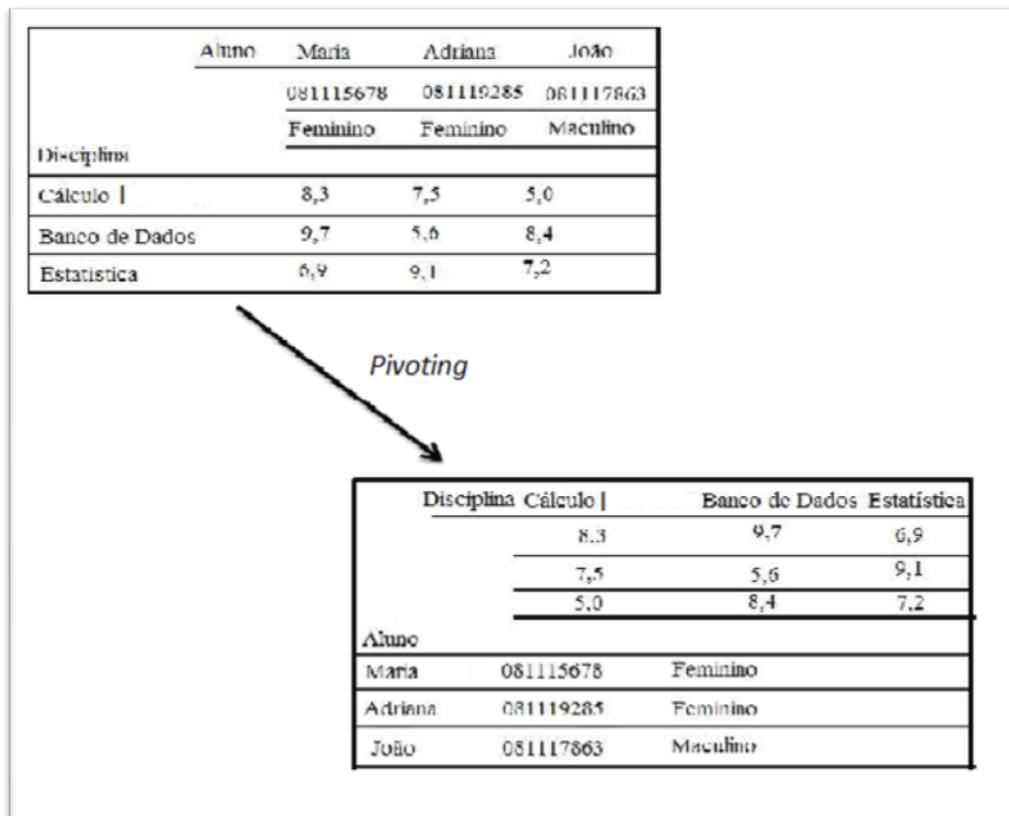


FIGURA 15: Exemplo de *Pivoting* em tabelas de dimensão

FONTE: Adaptado pelo próprio autor segundo Medeiros (2011)

2.3 CONSIDERAÇÕES FINAIS DO CAPÍTULO

Ao longo deste capítulo foram abordados os principais conceitos sobre *Business Intelligence* e suas tecnologias *Data Warehouse* e OLAP. Como visto, estas práticas são de grande importância para o processo de decisões dentro de um ambiente empresarial. No próximo capítulo serão aplicados conceitos de BI em um estudo de caso na UFPB, com uma proposta de aplicação da tecnologia *Data Warehouse* na STI.

3 ELABORAÇÃO DO PROTÓTIPO DE UM DATA WAREHOUSE

Neste capítulo serão aplicados os conceitos de implantação de uma solução BI para a Superintendência de Tecnologia da Informação da UFPB utilizando a ferramenta *QlikView Personal* 10 para demonstrar a criação desta solução.

3.1 A EMPRESA

A Superintendência de Tecnologia da Informação (STI) é um órgão da UFPB responsável pela infraestrutura de *software* e *hardware* dentro da universidade, além de oferecer suporte no uso de tecnologia da informação e comunicação aos três seguimentos: servidores, professores e estudantes. Sua importância para a instituição atualmente é igualada a de uma pró-reitoria. A STI gerencia alguns sistemas integrados, de caráter administrativo e acadêmico, assim como suas respectivas bases de dados.

Uma das bases de dados gerenciadas pela Superintendência é a do Sistema de Controle Acadêmico – SCA, nela estão armazenados todos os dados referentes ao meio acadêmico da UFPB, e o SGBD utilizado para guardar esta base é o *Oracle*². Esta base possui uma grande quantidade de dados armazenados, é composta por 87 tabelas e algumas delas possuem muitos campos. A tabela Alunos, por exemplo, possui 76 campos, onde estes armazenam dados de mais de 30 mil alunos que possuem cadastro na base, isto pode gerar problemas ao realizar consultas, devido à vasta quantidade de dados.

3.2 PROBLEMÁTICA

A STI recebe frequentemente uma grande demanda de solicitações de vários órgãos da UFPB, a respeito de relatórios para análises de informações de diversos interesses e assuntos. Estes relatórios são gerados pela gerência do setor de Inteligência Institucional através de planilhas Excel para onde os dados armazenados no BD operacional são importados, e também, por meio de documentos do Word.

Na base do SCA, onde são realizadas as buscas a dados referentes ao meio acadêmico da universidade, é armazenada uma imensa quantidade de dados dentro de *views* e tabelas sem normalização, com isso, a filtragem dos dados em consultas que envolvam várias tabelas, por exemplo, pode levar muito tempo e não ser feita com eficiência.

² <http://www.oracle.com/us/corporate/history/index.html>

Além disso, análises de relatórios gerados através de consultas a dados em banco de dados operacionais, em alguns casos levam a uma difícil compreensão. Por exemplo, um assunto bastante discutido em reuniões entre chefes de departamentos e cursos, é o índice de evasão dos alunos dentro da UFPB, e também a maneira de fazer este cálculo, pois esta evasão pode ser interpretada de diversas formas. Em consultas a base operacional SCA, este índice não é mostrado de uma forma clara, dificultando assim para os gestores, o processo de tomada de decisões.

Uma tecnologia de *Data Warehouse*, onde os dados nela carregados possam ser originados de diversas fontes, que realize buscas de forma mais rápida e que permita uma melhor compreensão das informações nas análises dos relatórios gerados, seria de grande valia para a STI auxiliar no processo de tomada de decisões dos gestores dos órgãos correspondentes a UFPB.

3.3 APLICANDO DATA WAREHOUSE NA STI

Com o intuito de facilitar o processo de tomada de decisões dos gestores dos órgãos da UFPB, e auxiliar a STI com as demandas referentes a relatórios gerados através dos dados exportados da base do Sistema de Controle Acadêmico, foi proposta a aplicação de *Data Warehouse* no ambiente de BI da STI. Para o desenvolvimento desta aplicação neste trabalho, serão analisados os seguintes casos dentro do curso de Sistemas de Informação – Campus IV – Litoral Norte:

- Índice de evasão dos alunos, por período;
- Índice de reprovação dos alunos nas disciplinas.

3.3.1 Banco de Dados do Sistema de Controle Acadêmico

Como visto no tópico anterior, os dados para as análises feitas neste trabalho são exportados do banco de dados do Sistema de Controle Acadêmico (SCA), que guarda todos os dados referentes ao meio acadêmico da Universidade Federal da Paraíba.

A base SCA apresenta tabelas sem normalização, além disso, alguns dados são representados nas tabelas por números, e não por descrições. Para representar o estado civil,

por exemplo, ao invés de descrições do tipo ‘solteiro’, ‘casado’, na base estes são representados por ‘1’, ‘2’, dificultando assim a interpretação destes dados.

A Tabela 01 apresenta a quantidade de campos e o número de registros que constam nas respectivas tabelas (Alunos, Cursos, Cursos Turma, Disciplinas, Matrículas) que terão seus dados exportados para análise neste trabalho, advindos da base de dados.

| Nome da Tabela | Quantidade de Campos | Quantidade de Registros |
|----------------|----------------------|-------------------------|
| Alunos | 79 | 190.368 |
| Cursos | 24 | 247 |
| Matrículas | 14 | 6.873.557 |
| Disciplinas | 19 | 12.819 |

TABELA 01: Informações sobre tabelas do SCA

Os dados destas tabelas serão filtrados e exportados para a ferramenta *QlikView*. A mesma será apresentada a seguir.

3.3.2 *QlikView*³

QlikView é uma ferramenta de BI desenvolvida pela *QlikTech*⁴ que combina interatividade e análise de informações em tempo real. Ela pode ser utilizada desde como uma ferramenta de relatórios em formato de tabela até a criação de gráficos elaborados e painéis de controle (*dashboards*) complexos com dados advindos de uma enormidade e variedade de origens.

³ [QlikView Passo a Passo - SWB.docx](#)

⁴ A *QlikTech* sediada em *Radnor*, Pensilvânia, é uma empresa que trabalha com a tecnologia “*in-memory-AdHoc*” desde 1993 e hoje está apoiada no conceito de uso de base de dados não relacional, de alta compressão e totalmente residente em memória.

Este *software* é considerado como uma ferramenta que permite aos usuários trabalhar facilmente com dados originados de diversas fontes, como por exemplo: bases de dados (SQL Server ou Oracle) arquivos do Excel, ou até documentos de texto.

O *QlikView* possui uma variedade de opções para análises. Gráficos, diagramas e tabelas podem ser utilizados para apresentar os dados de diversas formas.

Segundo a documentação em Português do *QlikView*, para *desktop* a ferramenta pode ser usada gratuitamente, como a versão *Personal Edition*. Com o *QlikView Personal Edition*, pode-se usar integralmente as funções da ferramenta, mas não é possível abrir documentos criados por outros usuários. Para isso, é necessária uma licença do *QlikView*.

Uns dos motivos pelo qual se escolheu o *QlikView* neste trabalho, é por esta ferramenta ser orientada ao usuário e permitir a elaboração de relatórios com facilidade. Após ser realizado o processo de ETL, os relatórios podem ser produzidos com a inclusão de elementos gráficos e conhecimentos básicos de funções de agregação, somas e contagens por exemplo. (DA COSTA, 2011).

Para o desenvolvimento deste trabalho, a versão do *QlikView* instalada foi a *QlikView Personal Edition 10*, no sistema operacional Windows 7 *Ultimate* e *Professional*, e os arquivos importados nela são do tipo .xls (planilhas do Excel).

3.3.3 Processo ETL

Para a aplicação do *Data Warehouse* foi necessário passar pelo processo de extração, transformação, e carga dos dados (ETL) das *views* da base SCA, depois desta etapa, é possível fazer as análises que o trabalho propõe.

Para a extração dos dados da base SCA do SGBD *Oracle*, foi preciso exportar arquivos no formato xls para o *QlikView*. Foram criados *scripts* de extração de dados, estes *scripts* são responsáveis por se conectar a base de dados e realizar a extração. Cada *script* é composto por comandos *select*, onde cada *select* gera uma tabela no arquivo criado no *QlikView*. Todos os *scripts* utilizados neste processo encontram-se no apêndice deste trabalho.

A Figura 16 mostra um exemplo de extração de dados da tabela Alunos do SCA para a ferramenta.

```

DimAluno :                               /* Nomeando a dimensão*/

SELECT `ALU_MATRICULA` AS ID_ALUNO,      /* Carga dos dados da Tabela Alunos*/
      `ALU_CCU_CUR_COD_CURSO`,
      `ALU_CCU_COD_CURRICULO`,
      `ALU_CAMPUS`,
      `ALU_NOME`,
      `ALU_SEXO`,
      `ALU_SITUACAO`,
      `ALU_FORMA_EVASAO`,
      `PER_EVASAO`,
      `ALU_FORMA_INGRESSO`,
      `ALU_DT_CONCLUSAO`,
      `ALU_DT_EXP_DIPLOMA`,
      `PER_INGRESSO`,
      `ALU_DT_ENC`,
      `ALU_DEFICIENCIA`,
      `E_COTISTA`,
      `ACAO_AFIRMATIVA_ID`

FROM `C:\Users\Giordania\AlunosSistemas.xls`.`Exportar Planilha$`;

```

FIGURA 16: Extração dos dados da tabela Alunos no *QlikView*

É importante ressaltar que esta extração foi realizada para todas as demais tabelas envolvidas neste trabalho.

Após a etapa de extração, inicia-se a etapa de transformação dos dados, onde os mesmos são transformados para que haja relação entre eles nas tabelas dimensionais. Para melhor entendimento, a Figura 17 mostra um exemplo de transformação de dados na tabela Matrículas através de comandos SQL. As transformações dos dados são feitas para cada carga. No caso da tabela Matrículas, as transformações foram feitas respectivamente nas linhas 25 e 32. Na primeira criou-se uma chave primária composta por 04 campos da tabela e foi nomeada de ID_MATRICULA, e na segunda, o campo MAT_TUR_DIS_DISCIPLINA foi nomeado de ID_DISCIPLINA para poder se relacionar com a dimensão Disciplina no modelo dimensional. Todas as transformações são feitas dentro do comando LOAD.

```
DimMatricula : /* Nomeando as Dimensões*/

/*Atribuindo (MAT_TUR_TURMA&MAT_TUR_DIS_DISCIPLINA&MAT_TUR_ANO&MAT_TUR_PERIODO) para ID_MATRICULA como chave primária*/
LOAD AutoNumberHash256 (MAT_TUR_TURMA&MAT_TUR_DIS_DISCIPLINA&MAT_TUR_ANO&MAT_TUR_PERIODO) AS ID_MATRICULA,

'MAT_ALU_MATRICULA',
'MAT_TUR_TURMA',
'MAT_TUR_DIS_DISCIPLINA' AS ID_DISCIPLINA, /* Renomeando MAT_TUR_DIS_DISCIPLINA para ID_DISCIPLINA para Matriculas
                                             se ligar com a tabela Disciplinas*/
'MAT_TUR_ANO',
'MAT_TUR_PERIODO',
'MAT_SITUACAO',
'MAT_MEDIA_FINAL'

FROM 'C:\Users\Giordania\matriculassialunos.xls'.'Exportar Planilha$!';
```

FIGURA 17: Transformação dos dados na tabela Matrículas

Cada tabela dimensional foi construída com dados extraídos da base original, possuindo seus próprios atributos e suas respectivas chaves de identificação. Estes atributos serão utilizados como seletores nas buscas que serão feitas para as análises.

A tabela de Fatos é composta pelas chaves primárias das tabelas dimensionais e pela lógica da ligação entre elas. A Figura 18 mostra que todas as tabelas de dimensões são ligadas através de suas chaves primárias. Percebe-se que a tabela de dimensão Matricula se relaciona com todas as outras tabelas, ou seja, em seu comando LOAD além de sua chave primária, possui também as chaves primárias das dimensões Alunos e Cursos.

```
FatosAcadêmico :      /* Nomeando a tabela de Fatos*/

LOAD `CUR_COD_CURSO` AS ID_CURSO      /*Carregando a tabela através da chave primária */

FROM `C:\Users\Giordania\cursoSI.xls`.`Exportar Planilha$`;

Concatenate      /* Fazendo a junção das tabelas */

LOAD `ALU_MATRICULA` AS ID_ALUNO,
     `CUR_COD_CURSO` AS ID_CURSO

FROM `C:\Users\Giordania\AlunosSistemas.xls`.`Exportar Planilha$`;

Concatenate

LOAD AutoNumberHash256 (MAT_TUR_TURMA&MAT_TUR_DIS_DISCIPLINA&MAT_TUR_ANO&MAT_TUR_PERIODO) AS ID_MATRICULA,
     `ALU_MATRICULA` AS ID_ALUNO,
     `CUR_COD_CURSO` AS ID_CURSO

FROM `C:\Users\Giordania\matriculassialunos.xls`.`Exportar Planilha$`;
```

FIGURA 18: Criação da tabela de Fatos

Após a extração e transformação, é realizada a carga dos dados, onde os mesmos já estão transformados e prontos para a construção dos cenários para as análises.

Para todo o processo, foram escolhidas as tabelas Alunos, Matrículas, Cursos, e Disciplinas, pois, as análises realizadas neste trabalho envolverão assuntos referentes a alunos e disciplinas de um determinado curso, especificamente, o curso de Sistemas de Informação.

A Figura 19 apresenta o modelo Floco de Neve do projeto, composta por uma tabela de Fatos que foi nomeada de “FatosAcadêmico”, e três tabelas dimensionais: “DimAluno”, “DimCurso”, e “DimMatriculas”, ligadas a ela através de suas chaves primárias. Além disso, existe outra tabela dimensional nomeada de “DimDisciplina” que se liga apenas com a tabela dimensional “DimMatricula”. Esta ligação é feita através do campo “ID_DISCIPLINA”.

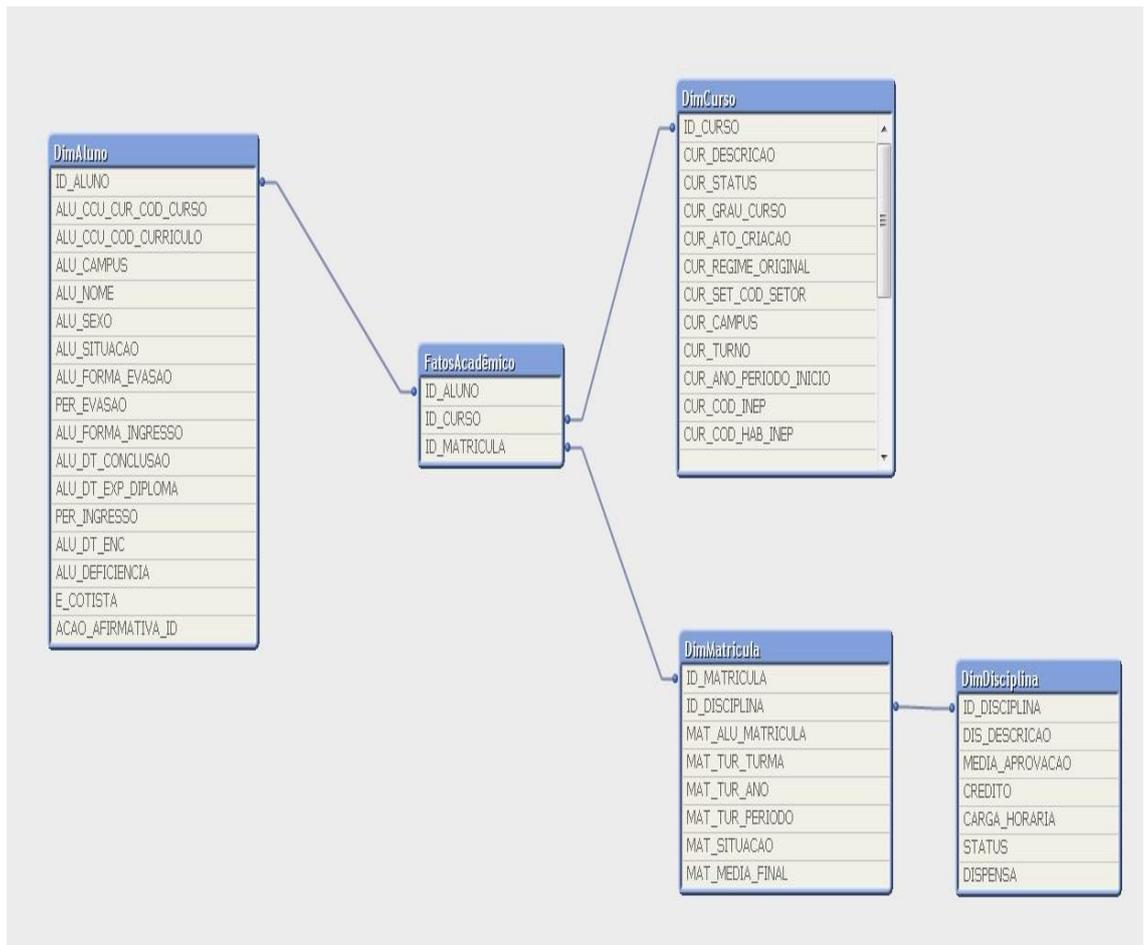


FIGURA 19: Modelo dimensional Floco de Neve gerado pelo *QlikView*

Com a visualização do modelo dimensional, é possível observar as ligações entre as tabelas através de suas chaves primárias na tabela de Fatos, como também os demais atributos.

Todo o processo de ETL e criação do DW através dos dados originados da base SCA podem ser representados através da Figura 20, que mostra com detalhes a estrutura de todo o processo.

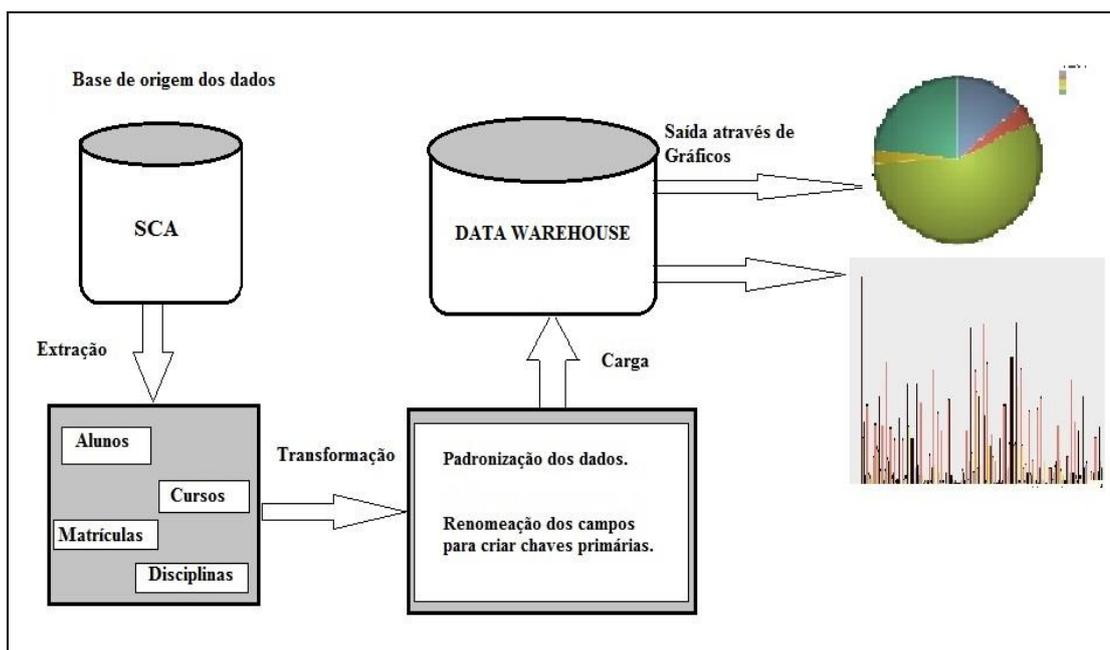


FIGURA 20: Processo de ETL e criação do DW através dos dados da base SCA

3.3.4 Construindo Cenários

Para analisar o índice de evasão e as disciplinas com maior índice de reprovação, foram criados alguns painéis que mostrarão de diferentes formas as informações das análises propostas. Estes cenários utilizam gráficos do tipo barra e pizza, e seletores como um objeto de Lista, por exemplo, para apresentar as informações ao usuário de uma forma mais clara e detalhada.

O primeiro painel, sobre os índices de reprovação nas disciplinas, mostra uma visão geral da situação dos alunos matriculados nas disciplinas do curso, onde é mostrado o índice dos alunos que foram aprovados, reprovados por média, reprovados por falta, ou realizaram trancamento, para cada disciplina do curso entre os anos de 2007 e 2013. A Figura 21 apresenta essa visão através de um gráfico de barras utilizando a dimensão Disciplina. Como exemplo para visualização dos índices de reprovação, foi escolhida a disciplina de Introdução a Programação.

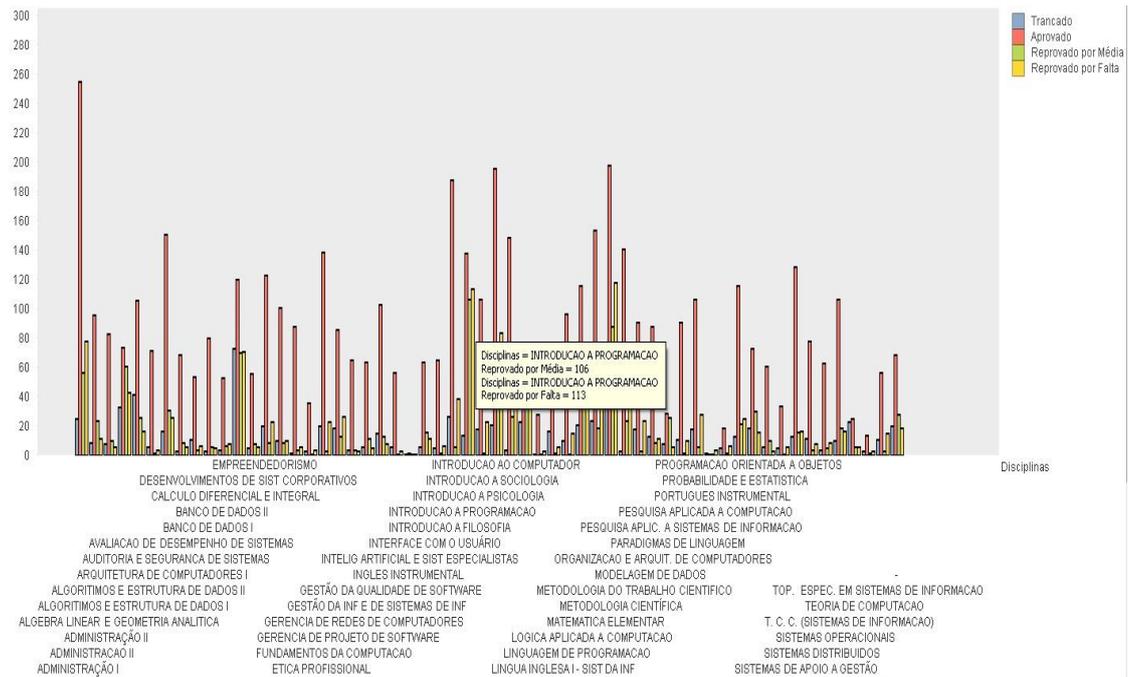


FIGURA 21: Visão geral da situação dos alunos matriculados nas disciplinas do curso de SI

A figura acima mostra uma visão geral da situação dos alunos matriculados nas disciplinas do curso de SI. Além disso, destaca o número de alunos que foram reprovados por média e reprovados por falta durante os anos de 2008 a 2013, na Disciplina de Introdução a Programação.

Para obter uma visão mais detalhada da situação dos alunos nas disciplinas, foram utilizados os seletores Disciplinas e Ano de Matrícula através de uma Tabela Dinâmica que além destes, contém também a Quantidade de Alunos Matriculados durante os respectivos anos. Com isso, pode-se comparar a quantidade de alunos matriculados nos anos com a quantidade de alunos reprovados por média, reprovados por falta, aprovados, e o total de trancamentos. A Figura 22 mostra esta visão na Disciplina de Introdução a Programação, através do gráfico de barras.

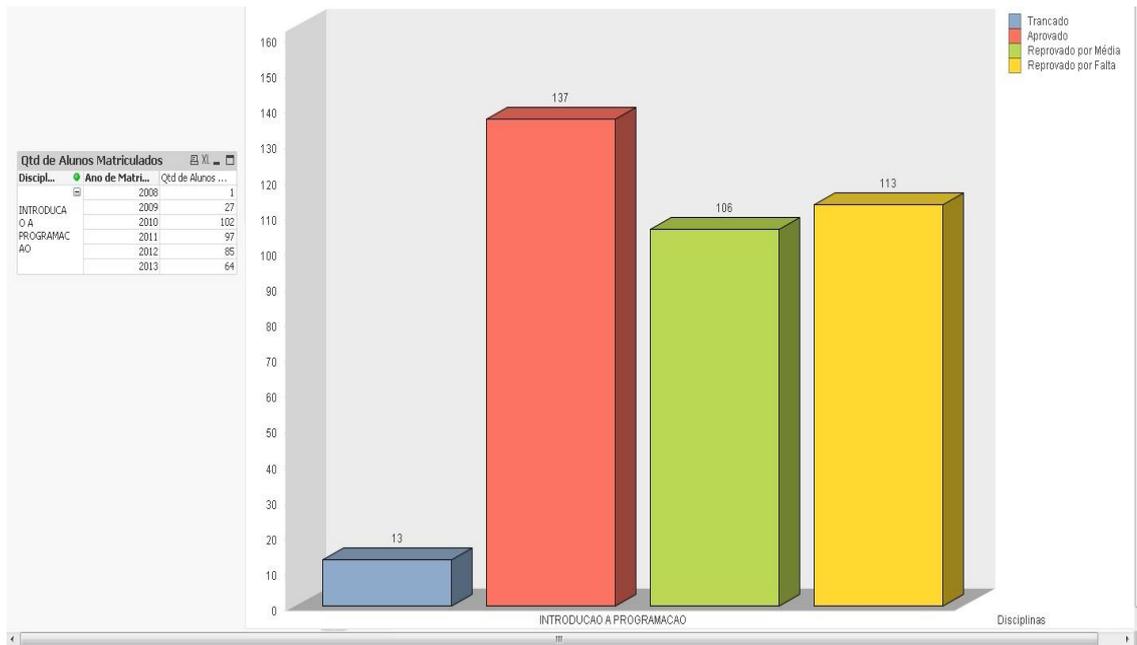


FIGURA 22: Índice de reprovação na disciplina Introdução a Programação durante os anos de 2008 a 2013

De acordo com as informações apresentadas no gráfico, pode-se observar que os alunos que reprovaram a disciplina por média durante os anos, foram 106 de 376 matriculados, e 113 foram reprovados por falta.

A Figura 24 apresenta o mesmo cenário da Figura 23, desta vez com a disciplina de Cálculo Diferencial e Integral entre os anos de 2007 a 2013.

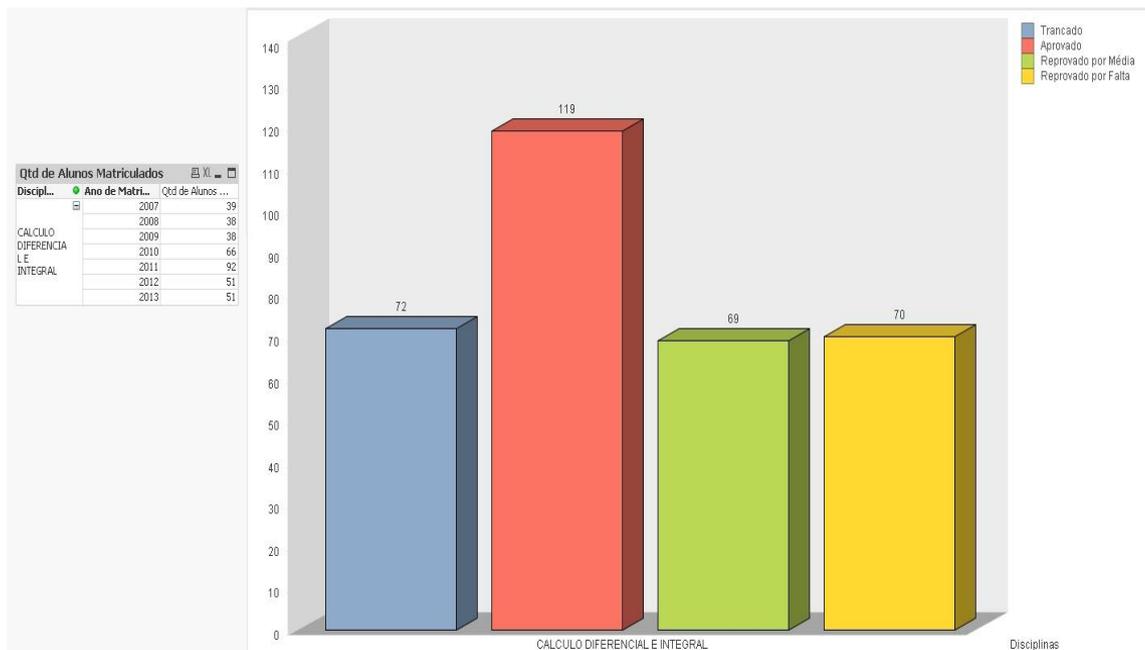


FIGURA 23: Índice de reprovação na disciplina Cálculo Diferencial a Integral durante os anos de 2007 a 2013

De acordo com as informações apresentadas no gráfico, pode-se observar que os índices de alunos trancados, reprovados por média, e reprovados por falta são praticamente os mesmos. Os alunos que trancaram a disciplina durante os anos são 72 de 375 matriculados, e 69 dos que continuaram na disciplina foram reprovados por média.

Através da Tabela Dinâmica é possível fazer a operação *Slice* nas informações, ou seja, extrair um nível de dados para obter informações apenas deste. A Figura 24 mostra a situação dos alunos na disciplina Introdução a Programação apenas no ano de 2013, onde o total de Alunos matriculados é de 64.

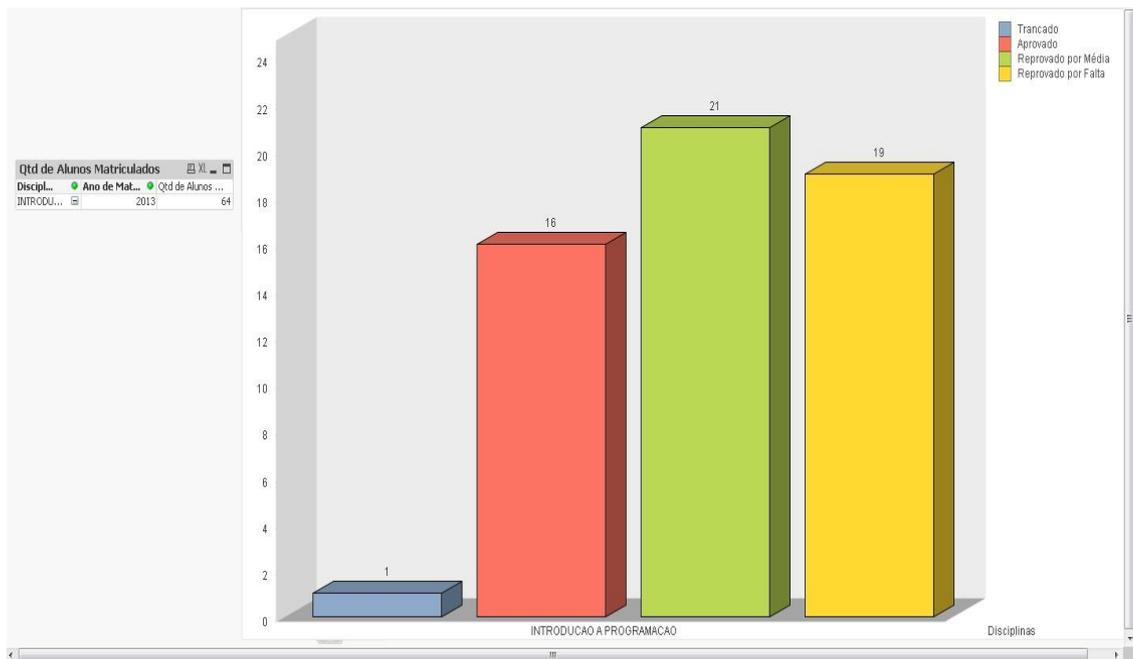


FIGURA 24: Índice de reprovação da disciplina Introdução a Programação no ano de 2013

De acordo com o gráfico acima, nota-se que ano de 2013 a quantidade de alunos reprovados por média é de 21 de 64 alunos matriculados, ou seja, aproximadamente 33% dos alunos que cursaram a disciplina em 2013 foram reprovados por média. É importante ressaltar que 07 dos 64 alunos matriculados ainda estão em curso.

Outra visão que se pode ter sobre as Disciplinas do curso de SI é a quantidade de alunos matriculados em um determinado ano ou período. Para esta análise foi utilizado o Objeto Lista que também contém seletores que podem mostrar a variação das informações em um gráfico. A Figura 25 mostra a quantidade de alunos matriculados nas disciplinas do curso de Sistemas de Informação no ano de 2013.

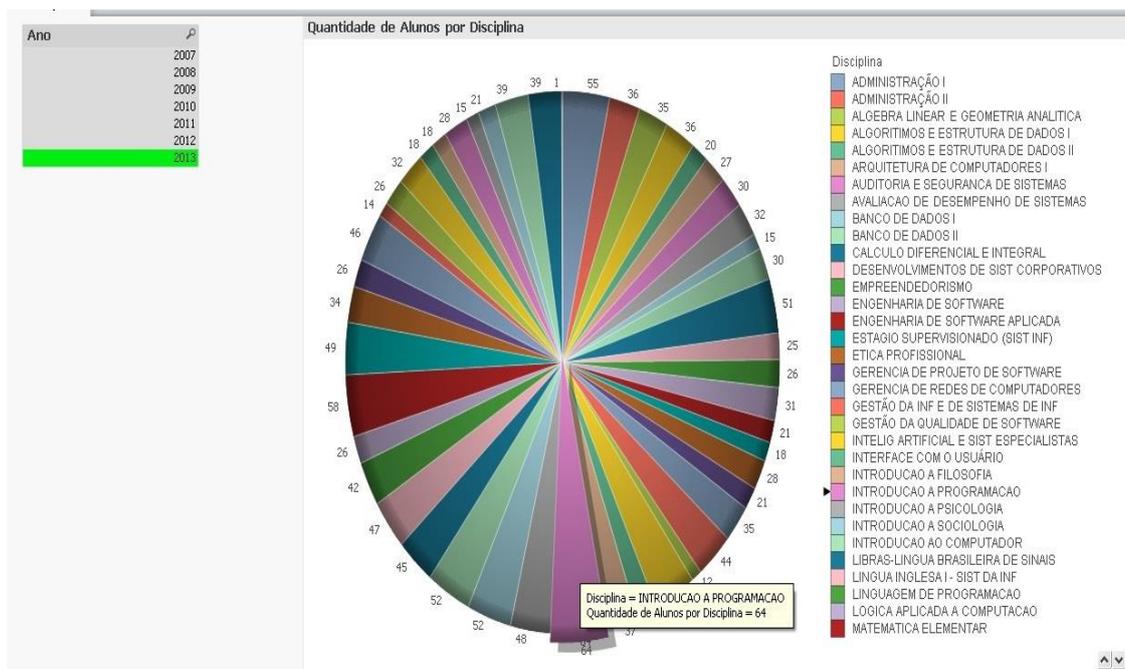


FIGURA 25: Quantidade de alunos matriculados nas disciplinas no ano de 2013

De acordo com o gráfico pode-se observar que a disciplina com mais alunos matriculados no ano de 2013 foi Introdução a Programação, com o total de 64 alunos.

Para a STI, alunos evadidos significam alunos que saíram do curso, seja por conclusão ou algum outro fator, como por exemplo: abandono, falecimento, transferência, mudança de curso, ou solicitação do aluno. O segundo painel, Índice de Evasão, mostra uma visão geral dos alunos evadidos do curso de SI entre os períodos de 2008.1 a 2013.2, para esta análise foi selecionada a Dimensão Alunos, como também as formas de evasão. A Figura 26 apresenta esta visão através de um Gráfico de Pizza, no painel também se encontra um objeto Lista para selecionar os períodos e mostrar a variação da evasão entre eles, e também uma caixa de texto com a legenda das formas de evasão que no gráfico são representadas por números. Além disso, o gráfico também apresenta a quantidade de alunos que não evadiram do curso.

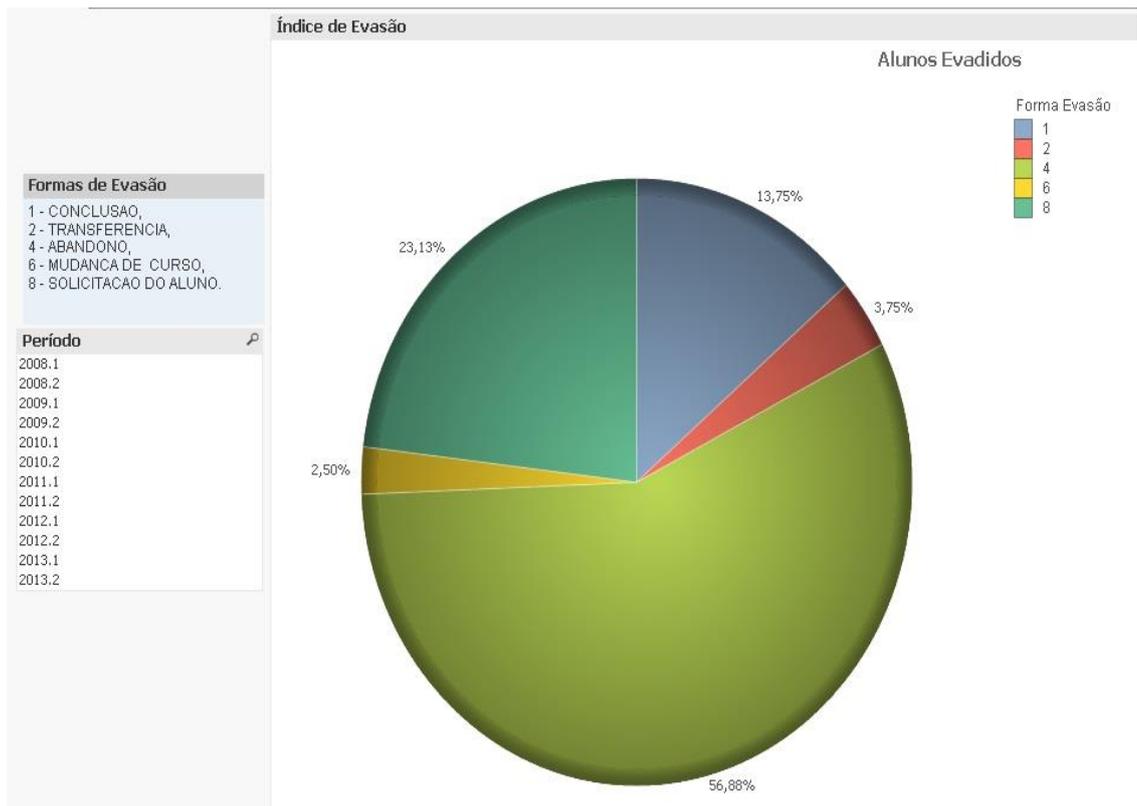


FIGURA 26: Visão geral dos alunos evadidos do curso entre os períodos 2008.1 a 2013.2

De acordo com o gráfico, dentre as formas de evasão, o maior índice é o de alunos que evadiram do curso por abandono, com aproximadamente 57 % dos alunos.

Para obter uma análise mais detalhada dessa visão, foi utilizado o seletor Período para selecionar o período pelo qual se deseja analisar. A Figura 27 mostra o índice de evasão dos alunos no período 2010.2 através do Gráfico Pizza.

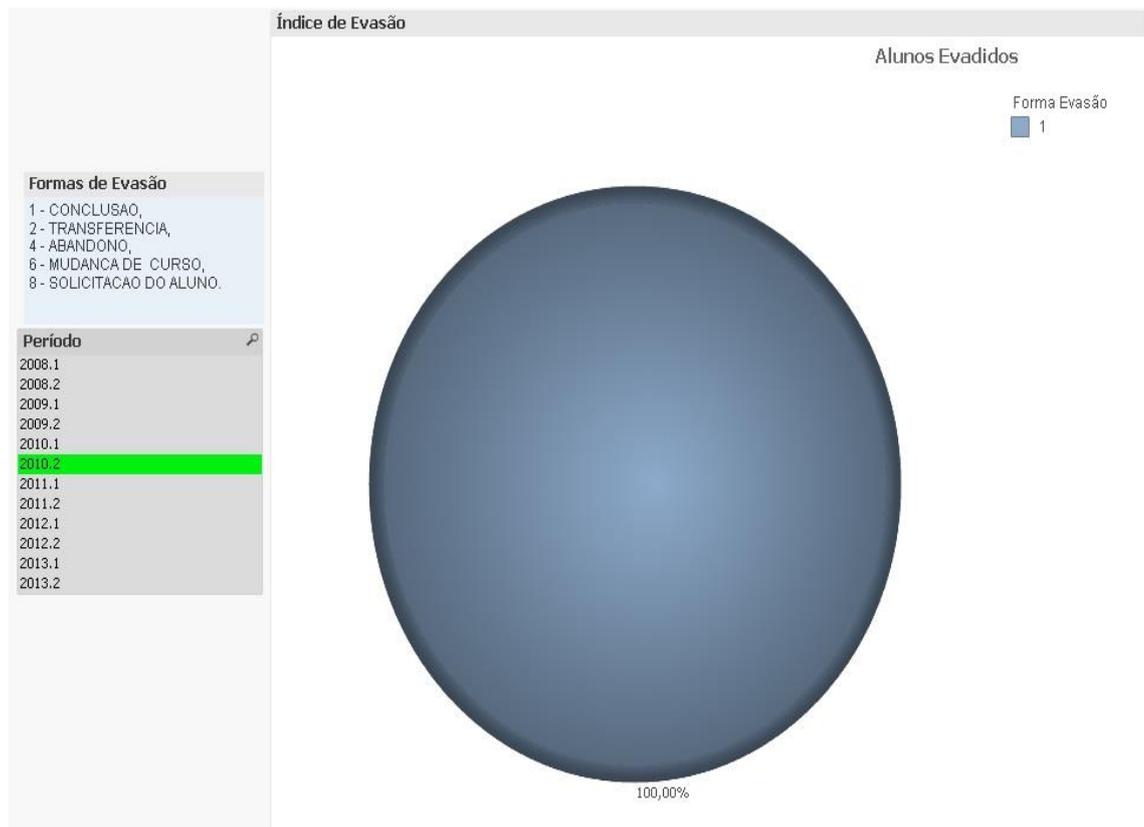


FIGURA 27: Quantidade de alunos evadidos no período 2010.2

É interessante observar que neste período só houve 01 aluno evadido no curso, onde o mesmo evadiu por forma de conclusão.

Utilizando o mesmo seletor da Figura 27, a Figura 28 mostra a quantidade de alunos evadidos, desta vez no período de 2011.1.

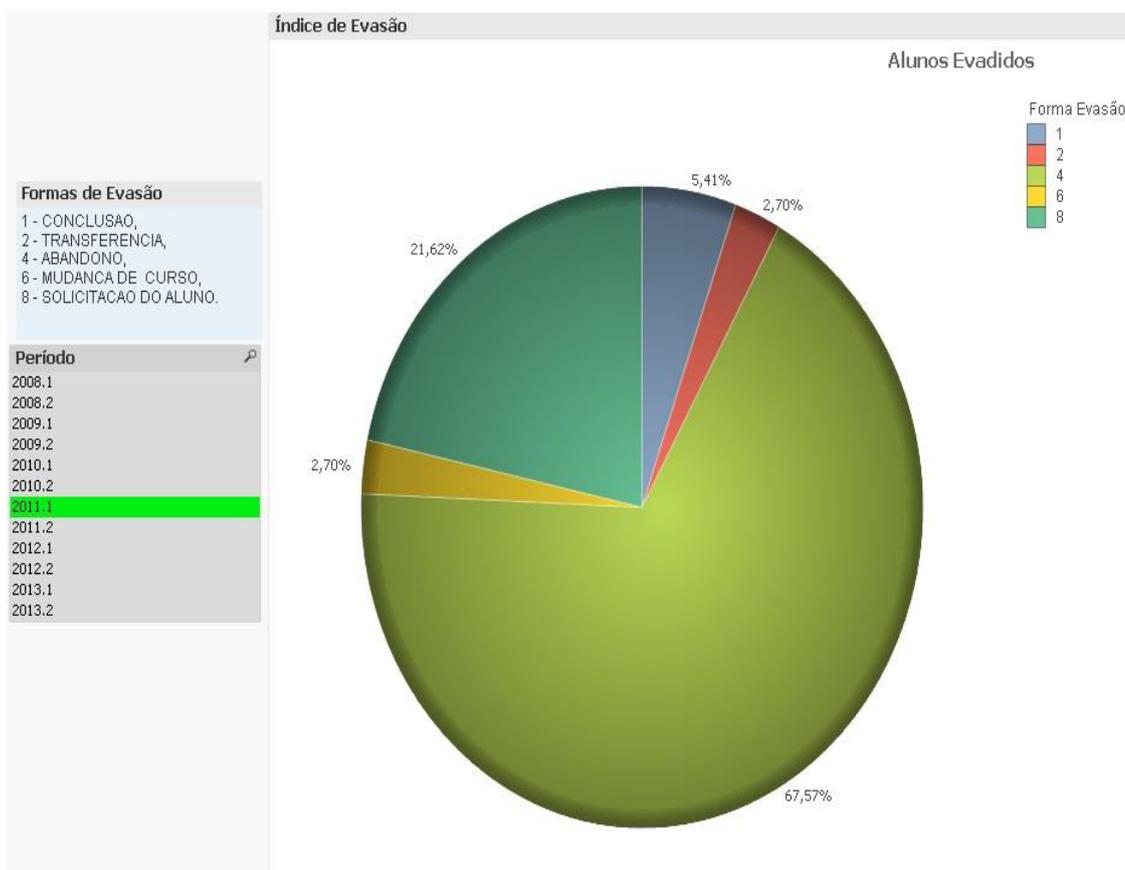


FIGURA 28: Quantidade de alunos evadidos no período 2011.1

É importante analisar que, ao contrário do gráfico da figura anterior onde só houve uma forma de evasão, e a quantidade de apenas 01 aluno evadido, este gráfico mostra uma grande variação dentre as formas de evasão, e também, uma grande quantidade de alunos evadidos por abandono. Aproximadamente 68% dos alunos evadidos neste período, evadiram por esta forma de evasão.

Outras formas de apresentação do índice de evasão são possíveis. A Figura 29 apresenta uma visão geral do índice de evasão durante os períodos de 2008.1 a 2013.2 através de um Gráfico de Barras.

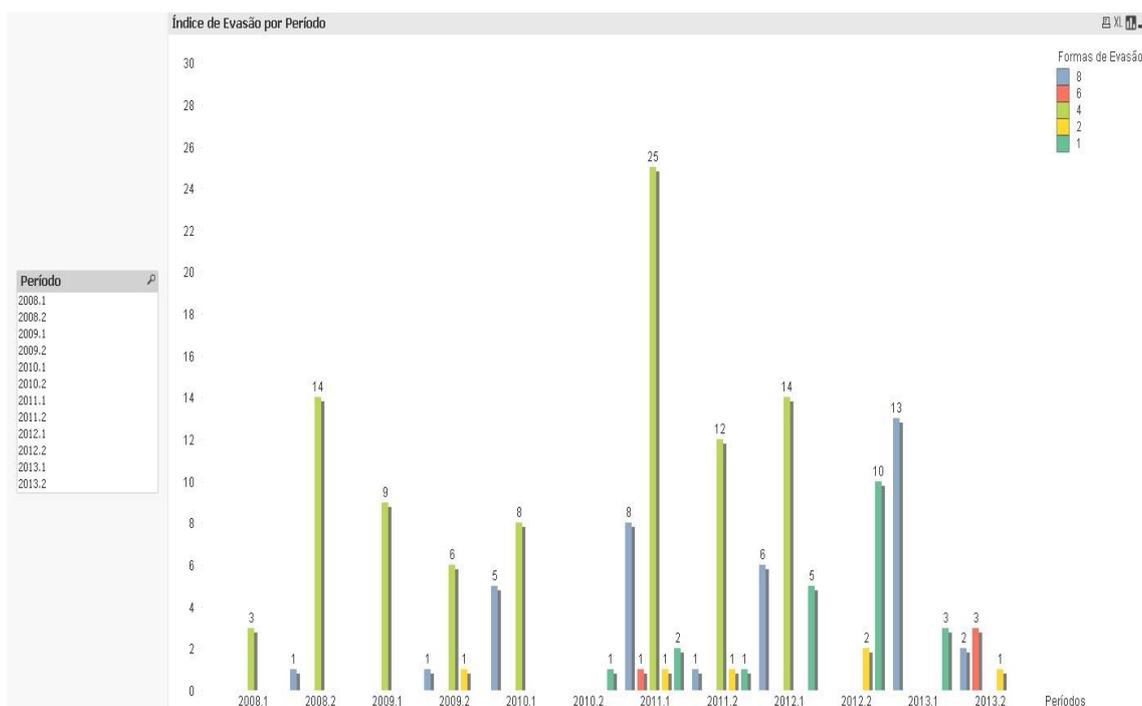


FIGURA 29: Visão geral dos alunos evadidos do curso de SI entre os períodos 2008.1 a 2013.2

Diante deste cenário pode-se observar que, dentre os alunos que evadiram durante os períodos, a maior representação é a da evasão por abandono. A evasão por abandono está representada no gráfico pela cor verde.

3.4 CONSIDERAÇÕES FINAIS DO CAPÍTULO

Este capítulo apresentou o desenvolvimento da proposta de implementação de *Data Warehouse* na STI com base nos conceitos definidos no capítulo 2. Para mostrar as análises, foram feitos cenários com alguns gráficos, e as respectivas interpretações dos mesmos. Através deles puderam-se obter informações de uma forma mais clara, e principalmente, uma dinamicidade na busca das informações. Outras análises podem ser feitas pelos tomadores de decisão, para isso, basta solicitar mudanças nos seletores, adicionar/alterar dimensões, e a ferramenta as realizam de forma hábil.

O próximo capítulo apresenta as conclusões deste trabalho e algumas propostas para possíveis trabalhos futuros.

4 CONCLUSÃO

Com este trabalho pode-se concluir o quão importante é a utilização de uma ferramenta de BI para o processo de tomada de decisões dos gestores dentro de uma organização, pois, através dos gráficos gerados pôde se obter uma visão mais clara das informações, possibilitando praticidade com relação às análises. No trabalho foi apresentada uma proposta de implantação de *Data Warehouse* no ambiente de *Business Intelligence* da STI.

Primeiramente foi realizado um estudo bibliográfico sobre os temas abordados, e com base nesse estudo, foi feito o desenvolvimento do trabalho enfatizando a necessidade que a STI tem de otimizar a geração de relatórios com informações sobre os dados que a mesma gerencia. Para as análises, foram escolhidos temas considerados importantes para o processo de decisões de gestores da UFPB, como o índice de evasão dos alunos, e o índice de reprovação nas disciplinas. Os dados extraídos para estas análises foram apenas do curso de Sistemas de Informação. É importante ressaltar que as informações apresentadas nos cenários foram validadas por um funcionário do setor de Inteligência Institucional da empresa.

Existiram dificuldades na realização deste trabalho, uma delas foi o acesso limitado aos dados contidos nos campos das tabelas do SCA, onde não era possível ver todas as informações que os dados representavam, além disso, as tabelas não são normalizadas. Outra dificuldade foi o acesso restrito à base, ou seja, só foi possível acessar a base de dados do sistema dentro do próprio ambiente da STI. Tudo isso acarretou no fator tempo para o desenvolvimento, principalmente para o processo de ETL e elaboração dos gráficos para as análises.

Em 2011, Larissa Queiroz desenvolveu em seu trabalho de conclusão de curso uma correlação entre o Perfil Comportamental e o Desempenho Acadêmico dos Alunos do curso de Ciências da Computação do UNIPÊ, utilizando a ferramenta *Data Warehouse* no ambiente de BI. Este trabalho mostrou uma análise de informações extraídas de fontes diferentes, e auxiliou os gestores no processo de tomada de decisões.

É importante ressaltar que cada trabalho tem sua particularidade e seus objetivos. Na UFPB ainda não existe nenhum trabalho referente à *Data Warehouse*, esta prática de BI que permite uma melhor visualização das informações, auxiliando no processo de tomada de decisões. Uma proposta de aplicação de um *Data Warehouse* para auxiliar os gestores nos

processos de tomada de decisões correspondentes a comunidade acadêmica da UFPB, poderá ser um primeiro passo para a implantação de um *Data Warehouse* dentro da instituição.

A partir das análises feitas neste trabalho, os gestores poderão avaliar algumas questões referentes ao curso de Sistemas de Informação. Podem-se considerar propostas de análises para possíveis trabalhos futuros, como por exemplo: mostrar uma média de tempo que um aluno de SI leva para concluir o curso, analisar os motivos pelos quais levam os alunos a reprovarem as disciplinas, como também, sobre outros temas abordados, correspondentes ao meio acadêmico da UFPB.

REFERÊNCIAS BIBLIOGRÁFICAS

ANTONELLI, A. S. Conhecendo o Business Intelligence (BI). **Revista TECAP**, v.3, n.3, 2009 anual. Disponível em: < revistas.utfpr.edu.br/pb/index.php/CAP/article/view/933/544 >. Acesso em: 07 jun. 2014.

ARAUJO, T. M. E. et. al. **Um estudo sobre as ferramentas OLAP**. 24 set. 2007. Disponível em: < <http://www.devmedia.com.br/um-estudo-sobre-as-ferramentas-olap/6691>>. Acesso em: 29 jun. 2014.

CALDAS, Marcos Paulo Kohler et. al. **Aplicações sobre um Data Warehouse no ambiente das organizações e suas vantagens**. XIII SIMPEP – Bauru, 2006.

CRUZ, R. **Arquitetura de um Data Warehouse**. 9 fev. 2011. Disponível em: < <http://www.100loop.com/destaque/arquitetura-de-um-data-warehouse-parte-i/>> . Acesso em : 10 jun. 2014.

DE PAULA, M. L. W. **Extract, Transformation and Load (ETL) – Ferramentas BI**. 07 mai. 2012. Disponível em: < <http://www.devmedia.com.br/extract-transformation-and-load-etl-ferramentas-bi/24408>>. Acesso em: 01 jun. 2014.

DA COSTA, X. V. M. A. **BUSINESS INTELLIGENCE – BI**. Uma aplicação de Business Intelligence na Advocacia. João Pessoa, 2011. Monografia (Graduação em Ciências da Computação) – Centro Universitário de João Pessoa.

DI DOMENICO, A. J. **Definição de um ambiente Data Warehouse em uma instituição de ensino superior**. Florianópolis, 2001. (Pós Graduação em Engenharia de Produção) – Universidade Federal de Santa Catarina.

FELTES, H. L. **Desenvolvimento de uma solução de BI para o ERP SIGER**. Novo Hamburgo, 2010. (Graduação em Ciências da Computação) – Universidade Feevale.

GAMA E ABREU, G. S. F. Desmistificando o Conceito de ETL. **Revista de Sistemas de Informação**, n.2, jul/dez. 2008. Disponível em: <http://www.fsma.edu.br/si/Artigos/V2_Artigo1.pdf>. Acesso em: 07 Jun. 2014.

GOMES, B. A. M. **Modelação de um Data Warehouse para a Direção Geral do Tesouro e Finanças e implementação de um Data Mart para o processo de Gestão Patrimonial**.

Lisboa, 2010. (Mestrado em Estatística e Gestão da Informação) – Universidade Nova de Lisboa.

INMON, H. W. The Data Warehouse Environment. In: **Building the Data Warehouse**. 3d. Canada, John Wiley & Sons, Inc, 2002. 428p.

KAUPA, P. **Business Intelligence (BI)**. 26 out. 2008. Disponível em: <<http://www.devmedia.com.br/business-inteligence-bi/10770>>. Acesso em: 31 mai. 2014.

KEMEZINSKI, A. et. al. Como obter vantagem competitiva utilizando Business Intelligence?. **Revista Produção**, v.3, n.2, jun 2003. Disponível em: <www.producaoonline.inf.br>. Acesso em: 27 mai. 2014.

LIMA, L. A. C. **Modelagem Dimensional – Conceitos Básicos**. 7 jan. 2010. Disponível em: < <http://litolima.com/2010/01/07/modelagem-dimencional-conceitos-basicos/>>. Acesso em: 20 jun. 2014.

LORENA, N. H. R. **Utilização de técnicas de Business Intelligence para avaliar o processo seletivo FATEC- SJC**. São José dos Campos, 2011. (Técnico em Banco de Dados)- Faculdade de Tecnologia de São José dos Campos.

MEDEIROS, Q. N. L. **Business Intelligence - BI: Uma correlação entre o Perfil Comportamental e o Desempenho Acadêmico dos Alunos do Curso de Bacharelado em Ciência da Computação – UNIPÊ**. João Pessoa, 2011. Monografia (Graduação em Ciências da Computação) – Centro Universitário de João Pessoa.

NOVAIS, C. R. R. **Modelagem Dimensional**. São Paulo, 2012. (Técnico em Processamento de Dados)- Faculdade de Tecnologia de São Paulo.

OLIVEIRA, T. D.; PEREIRA, J. O. **Um estudo do Business Intelligence no ambiente empresarial**.

PALITOT, B. A. A. **Análise comparativa de Arquiteturas de Distribuição de Data Warehouse**. Recife, 2007. (Graduação em Ciências da Computação) – Universidade Federal de Pernambuco.

PRIMAK, V. F. Objetivos do B.I. In: **Decisões com BI (Business Intelligence)**. Ed. Ciência Moderna. 2008.

ROHDEN, B. R. **Banco de Dados: Relacional x Multidimensional**. 19 nov. 2009. Disponível em: <<http://pt.scribd.com/doc/22742853/Artigo-Banco-de-Dados-Relacional-vs-Multidimensional>>. Acesso em: 20 jun. 2014.

SEZÕES, C. et. al. **Business Intelligence**. Porto, Sociedade Portuguesa de Inovação, 2006. 160p.

TAVARES, O. J. E. **Processo ETL: O caso da Unitel + Telecomunicações**. Cabo Verde, 2013. (Graduação de Engenharia de Sistemas e Informática) – Universidade Jean Piaget de Cabo Verde.

TEIXEIRA, R. J. **Introdução à Tecnologia Data Warehouse**. 31 mar. 2013. Disponível em: <<http://www.devmedia.com.br/introducao-a-tecnologia-data-warehouse/27629>>. Acesso em: 10 jun. 2014.

APÊNDICE

SCRIPTS

```
DimAluno :                               /* Nomeando a dimensão*/

SELECT `ALU_MATRICULA` AS ID_ALUNO,      /* Carga dos dados da Tabela Alunos*/
`ALU_CCU_CUR_COD_CURSO`,
`ALU_CCU_COD_CURRICULO`,
`ALU_CAMPUS`,
`ALU_NOME`,
`ALU_SEXO`,
`ALU_SITUACAO`,
`ALU_FORMA_EVASAO`,
`PER_EVASAO`,
`ALU_FORMA_INGRESSO`,
`ALU_DT_CONCLUSAO`,
`ALU_DT_EXP_DIPLOMA`,
`PER_INGRESSO`,
`ALU_DT_ENC`,
`ALU_DEFICIENCIA`,
`E_COTISTA`,
`ACAO_AFIRMATIVA_ID`

FROM `C:\Users\Giordania\AlunosSistemas.xls`.`Exportar Planilha$`;
```

```
DimAluno :

LOAD `ALU_MATRICULA` AS ID_ALUNO, /* Atribuindo ALU_MATRICULA para ID_ALUNO como chave primária */
`ALU_CCU_CUR_COD_CURSO`,
`ALU_CCU_COD_CURRICULO`,
`ALU_CAMPUS`,
`ALU_NOME`,
`ALU_SEXO`,
`ALU_SITUACAO`,
`ALU_FORMA_EVASAO`,
`PER_EVASAO`,
`ALU_FORMA_INGRESSO`,
`ALU_DT_CONCLUSAO`,
`ALU_DT_EXP_DIPLOMA`,
`PER_INGRESSO`,
`ALU_DT_ENC`,
`ALU_DEFICIENCIA`,
`E_COTISTA`,
`ACAO_AFIRMATIVA_ID`;

SQL SELECT *
FROM `C:\Users\Giordania\AlunosSistemas.xls`.`Exportar Planilha$`;
```

```

78
79 DimDisciplina :
80
81
82
83 LOAD `DIS_DISCIPLINA` AS ID_DISCIPLINA, /* Atribuindo DIS_DISCIPLINA para ID_DISCIPLINA como chave primária*/
84
85     `DIS_DESCRICAO`,
86     `MEDIA_APROVACAO`,
87     CREDITO,
88     `CARGA_HORARIA`,
89     STATUS,
90     DISPENSA;
91
92 SQL SELECT *
93 FROM `C:\Users\Giordania\DisciplinasSistemasdeInf.xls`.`Exportar Planilha$`;
94
95
96
97
98

```

```

42
43 LOAD `CUR_COD_CURSO` as ID_CURSO, /* Atribuindo CUR_COD_CURSO para ID_CURSO como chave primária*/
44
45     `CUR_DESCRICAO`,
46     `CUR_STATUS`,
47     `CUR_GRAU_CURSO`,
48     `CUR_ATO_CRIACAO`,
49     `CUR_REGIME_ORIGINAL`,
50     `CUR_SET_COD_SETOR`,
51     `CUR_CAMPUS`,
52     `CUR_TURNO`,
53     `CUR_ANO_PERIODO_INICIO`,
54     `CUR_COD_INEP`,
55     `CUR_COD_HAB_INEP`,
56     `CUR_MODALIDADE`,
57     `CUR_TITULO`,
58     `CUR_NOME`;
59
60 SQL SELECT *
61 FROM `C:\Users\Giordania\cursoSI.xls`.`Exportar Planilha$`;
62
63

```

```

DimMatricula :      /* Nomeando as Dimensões*/

/*Atribuindo(MAT_TUR_TURMA&MAT_TUR_DIS_DISCIPLINA&MAT_TUR_ANO&MAT_TUR_PERIODO) para ID_MATRICULA como chave primária*/
LOAD AutoNumberHash256 (MAT_TUR_TURMA&MAT_TUR_DIS_DISCIPLINA&MAT_TUR_ANO&MAT_TUR_PERIODO) AS ID_MATRICULA,

    `MAT_ALU_MATRICULA`,
    `MAT_TUR_TURMA`,
    `MAT_TUR_DIS_DISCIPLINA` AS ID_DISCIPLINA, /* Renomeando MAT_TUR_DIS_DISCIPLINA para ID_DISCIPLINA para Matriculas
                                                se ligar com a tabela Disciplinas*/
    `MAT_TUR_ANO`,
    `MAT_TUR_PERIODO`,
    `MAT_SITUACAO`,
    `MAT_MEDIA_FINAL`

FROM `C:\Users\Giordania\matriculassialunos.xls`.`Exportar Planilha$`;

```

```

FatosAcadêmico :      /* Nomeando a tabela de Fatos*/

LOAD `CUR_COD_CURSO` AS ID_CURSO      /*Carregando a tabela através da chave primária */

FROM `C:\Users\Giordania\cursoSI.xls`.`Exportar Planilha$`;

Concatenate      /* Fazendo a junção das tabelas */

LOAD `ALU_MATRICULA` AS ID_ALUNO,
     `CUR_COD_CURSO` AS ID_CURSO

FROM `C:\Users\Giordania\AlunosSistemas.xls`.`Exportar Planilha$`;

Concatenate

LOAD AutoNumberHash256 (MAT_TUR_TURMA&MAT_TUR_DIS_DISCIPLINA&MAT_TUR_ANO&MAT_TUR_PERIODO) AS ID_MATRICULA,
     `ALU_MATRICULA` AS ID_ALUNO,
     `CUR_COD_CURSO` AS ID_CURSO

FROM `C:\Users\Giordania\matriculassialunos.xls`.`Exportar Planilha$`;

```

ANEXO

Documento de Autorização para utilização dos dados da base de Dados do Sistema de Controle Acadêmico

UNIVERSIDADE FEDERAL DA PARAÍBA CENTRO DE CIÊNCIAS APLICADAS A EDUCAÇÃO BACHARELADO EM SISTEMAS DE INFORMAÇÃO

Solicitação

Eu, Giordania Giselle G. da Silva, vem por meio desta, solicitar ao V. S. Superintendente a permissão para que eu possa desenvolver o meu trabalho de conclusão de curso utilizando a base de dados do Sistema de Controle Acadêmico (SCA) gerenciada pela Superintendência de Tecnologia de Informação. O trabalho tem como principal objetivo mostrar uma proposta de criação de um *Data Warehouse* para auxiliar o ambiente de *Business Intelligence* da STI a atender demandas advindas dos órgãos correspondentes da UFPB.

No trabalho, as análises propostas por mim e pela Profa. Msc. Renata Viegas de Figueiredo através de buscas pelos dados dos alunos do curso de Sistemas de Informação - Campus IV são: mostrar o índice de evasão dos alunos desde 2006 até o ano corrente, as disciplinas que possuem maior e menor índice de reprovação, e também, calcular a média do tempo em que o aluno leva para concluir o curso. Para isto, precisarei realizar as seguintes tarefas:

- Gerar consultas através da base do SCA e importar arquivos gerados destas consultas, no formato xls, em uma ferramenta chamada *QlikView* para poder transformar e carregar os dados originados da base, e criar um modelo dimensional;
- Gerar gráficos através da ferramenta, para analisar as informações obtidas através dos dados;
- Analisar resultados que poderão auxiliar os órgãos correspondentes a UFPB no processo de tomada de decisões.

É importante informar que o trabalho será acompanhado pelo servidor da gerência de Inteligência Institucional da STI, Evandro Farias Rocha, matrícula SIAPE: 41937, e que todos os dados de caráter pessoal serão omitidos.

João Pessoa, 20 de Agosto de 2014.


Giordania Giselle G. da Silva


Evandro Farias Rocha


Superintendente

QlikView é uma ferramenta de BI orientada ao usuário, que auxilia na tomada de decisões a partir de fontes diversas de conhecimento, dados, pessoas e ambiente.