

Leonardo Campos do Carmo

Um Modelo de BI para Análise de Desempenho dos Alunos em Cursos EAD

Salvador

2014

Leonardo Campos do Carmo

Um Modelo de BI para Análise de Desempenho dos Alunos em Cursos EAD

Monografia apresentada à Banca Examinadora como exigência parcial para obtenção do título de bacharel em Sistemas de Informação pela Universidade do Estado da Bahia.

Universidade do Estado da Bahia – UNEB
Departamento de Ciências Exatas e da Terra
Colegiado de Sistemas da Informação

Orientador: Eduardo Manuel de Freitas Jorge

Salvador

2014

Leonardo Campos do Carmo
Um Modelo de BI para Análise de Desempenho dos Alunos em Cursos EAD/
Leonardo Campos do Carmo. – Salvador, 2014-
45 p. : il. (algumas color.) ; 30 cm.

Orientador: Eduardo Manuel de Freitas Jorge

Monografia (Graduação) – Universidade do Estado da Bahia – UNEB
Departamento de Ciências Exatas e da Terra
Colegiado de Sistemas da Informação, 2014.

1. Business Intelligence. 2. Ensino à distância. I. Eduardo Jorge. II. Universidade do Estado da Bahia. III. Departamento de Ciências Exatas e da Terra. IV. Um Modelo de BI para Análise de Desempenho dos Alunos em Cursos EAD

CDU 02:141:005.7

Agradecimentos

Em primeiro lugar a Deus por ter me amparado em cada momento.

- Aos meus pais pelo apoio e confiança.
- Ao meu orientador pelos incentivos e orientações.
- Aos meus amigos, Thomas, Felipe, Raylan, Marcelo, Kal e Daniel, pelos incentivos e companheirismo durante todo o curso.

*“É preciso força pra sonhar e perceber
Que a estrada vai além do que se vê.”
(Marcelo Camelo)*

Resumo

O provimento de informações pelo Business Intelligence para auxiliar gestores já é uma realidade. Os bancos de dados operacionais camuflam muitas informações importantes que podem servir de indicadores para um melhor gerenciamento. No entanto, ainda é pouco utilizado em áreas como a da educação, principalmente na área da educação à distância, que tem a difícil missão de tornar possível a avaliação processual num curso que tem como principal característica a comunicação não presencial, feita por meio das mídias. Com o objetivo de auxiliar os professores na avaliação processual no ensino à distância que surgiu o desafio desta pesquisa de criar um modelo de Business Intelligence para análise de desempenho dos alunos de um curso EAD. Este modelo é constituído de uma arquitetura técnica, uma arquitetura de BI e algumas visões analíticas.

Palavras-chaves: Business Intelligence. Educação à Distância. Data Warehouse.

Abstract

The information provided by business intelligence(BI) to help managers is already a reality. The operational databases omit a lot of information that can serve as indicators for better management. However, BI still little used in areas such as education, especially in the area of distance education. This area has the difficult task of making the process of evaluation be possible in a course whose main feature is communication through a media. In order to assist teachers in the process evaluation in distance education that came the challenge of this research to create a model of Business Intelligence for analyzing student performance in a distance learning course. This model consists of a technical architecture, a BI architecture and some analytical views.

Key-words: Business Intelligence. Distance education. Data Warehouse.

Lista de ilustrações

Figura 1 – Dados, Informação e Conhecimento	14
Figura 2 – Arquitetura Business Intelligence	17
Figura 3 – Integração dos Dados	20
Figura 4 – Drill Up	24
Figura 5 – Drill Down	24
Figura 6 – EAP	27
Figura 7 – Metodologia	28
Figura 8 – Mapeamento dos cubos.	30
Figura 9 – Data Mart de Frequência	32
Figura 10 – Data Mart de Quiz	33
Figura 11 – Variação da quantidade de alunos que concluíram o quiz por mês	34
Figura 12 – Variação da quantidade de concluintes.	35
Figura 13 – Variação da quantidade de concluintes da primeira semana do mês.	35
Figura 14 – Percentual da quantidade de alunos que concluíram o quiz por mês	36
Figura 15 – Percentual da quantidade de concluintes.	36
Figura 16 – Quantidade de alunos que concluíram o quiz por curso	37
Figura 17 – Quantidade de alunos que concluíram o quiz por curso no ano.	37
Figura 18 – Quantidade de quiz que foram concluídos de cada aluno no ano e mês.	38
Figura 19 – Quantidade de dias sem acesso por ano	38
Figura 20 – Quantidade de dias sem acesso por ano.	39
Figura 21 – Quantidade de dias sem acesso por mês do primeiro semestre do ano.	39
Figura 22 – Quantidade de dias sem acesso por curso e pelo tempo.	40
Figura 23 – Quantidade de dias sem acesso por curso e por mês.	40
Figura 24 – Página de acompanhamento do quiz.	41
Figura 25 – Quantidade de dias sem acesso por disciplina.	42
Figura 26 – Página de acompanhamento de acesso.	42

Lista de abreviaturas e siglas

BI	Business Intelligence
EAD	Educação à Distância
DW	Data Warehouse
AVA	Ambiente Virtual de Aprendizagem
SQL	Structured Query Language
ETL	Extract, Transform and Load
OLAP	On-Line Analytic Processing

Sumário

1	INTRODUÇÃO	11
2	FUNDAMENTAÇÃO TEÓRICA	13
2.1	Sistemas de Informação	13
2.1.1	Dados, Informações e Conhecimento	13
2.2	Educação à Distância (EAD) e o Moodle	15
2.3	Business Intelligence	16
2.3.1	Arquitetura e componentes do BI	17
2.4	Data Warehouse	18
2.4.1	Características do Data Warehouse	19
2.4.2	Metodologias Data Warehouse	20
2.5	Modelagem Dimensional	21
2.5.1	Tabelas Fato	21
2.5.2	Tabelas Dimensão	21
2.6	<i>Extract, Transform and Load – ETL</i>	22
2.7	On-Line Analytic Processing - OLAP	23
3	MODELO DE BI PARA ANÁLISE DE DESEMPENHO DOS ALUNOS NO EAD	24
3.1	Desenvolvimento do Modelo	26
3.1.1	Planejamento do Modelo	26
3.1.2	Metodologia do Modelo	27
3.2	Arquitetura Técnica e do Projeto	29
3.2.1	Camada de fontes de dados	29
3.2.2	Extract, Transform and Load - ETL	29
3.2.3	Data Warehouse e Data Marts	29
3.2.4	Especificações das Visões Analíticas	30
3.3	Modelagem Dimensional	30
3.3.1	Data Mart Frequência	31
3.3.2	Data Mart Quiz	32
3.4	Especificação das Visões Analíticas	33
3.4.1	Variação da quantidade de alunos que concluíram o <i>quiz</i> por mês	34
3.4.2	Percentual de alunos que concluíram o <i>quiz</i> no mês	36
3.4.3	Quantidade de alunos que concluíram o <i>quiz</i> por curso	36
3.4.4	Quantidade de dias que os alunos ficaram sem acessar o portal por Tempo	38
3.4.5	Quantidade de dias que os alunos ficaram sem acessar o portal por curso	39
3.5	Validação e Testes	40

4	CONSIDERAÇÕES FINAIS	43
	Referências	45

1 Introdução

A informação é de fundamental importância para o processo decisório de qualquer empresa. Com o mundo cada vez mais caracterizado com a rapidez das transformações nos diversos cenários, a informação assume um papel fundamental para tomada de decisões. A partir da década de 90 começa a revolução dos dados e da informação, com o apoio da internet, que se tornou democrática, invasiva e de amplo alcance. Os dados e as informações deixam de ser meros coadjuvantes. As empresas começam a ser induzidas a oferecer aos seus dirigentes melhores informações como forma de preparo para a competitividade do mercado, chega o momento da fidelização, da customização e da inteligência aplicada aos negócios (BARBIERI, 2011).

Os volumes de dados passaram a crescer, a internet cada vez mais presente no cotidiano, banco de dados organizacionais em constante crescimento. Esse acúmulo de dados foi se tornando complexo, de difícil extração e, conseqüentemente, de pouco valor para apoiar a tomada de decisão. Neste cenário, de desenfreado acúmulo de dados, surge o conceito de BI (*Business Intelligence*). Para suprir essa demanda, o BI surgiu como uma forma de extrair informações de um aglomerado de dados, antes impossível de se compreender. Essas informações serão a base para tomada de decisão nas organizações. O BI tem como principal pilar o *Data Warehouse*, onde ficam armazenados os dados a serem analisados. A partir da união de dados de diferentes fontes, que se analisados individualmente não trazem informações, surgem as informações estruturadas.

O BI esta cada vez mais presente em diversas áreas, como no segmento de saúde pública, segmento de transportes, de vendas em geral, etc. Dentre as diversas áreas de aplicabilidade do BI, o ramo da educação é de fundamental importância para auxiliar nos métodos de avaliação aplicados aos alunos, visto que, uma grande quantidade de dados são gerados todos os anos. Dados esses que são de difícil entendimento.

Um novo ramo da educação vem sendo colocado em evidência, a educação à distância. Que nada mais é do que a educação feita por meio de uma comunicação não presencial, feita por meio das mídias (vídeos, áudios, etc.). A educação à distância (EAD) rompe a relação espaço/tempo e se adequa ao cotidiano das pessoas.

Para acompanhar o processo de aprendizagem dos alunos, é necessário avaliá-los constantemente. Levantar o conhecimento prévio de cada aluno no decorrer das aulas conforme são aplicados os procedimentos de ensino. A esse acompanhamento contínuo da aprendizagem dos alunos dá-se o nome de Avaliação Processual. Diferentemente de uma situação de aprendizagem presencial, onde o professor tem contato direto diariamente com o aluno, tornando possível uma avaliação processual, na modalidade à distância essa

avaliação se torna um desafio. Uma forma de acompanhamento desses alunos é a análise de uso do Ambiente Virtual de Aprendizagem (AVA), ambiente esse onde são publicados trabalhos, atividades, vídeo-aulas, etc. Por exemplo, a informação de que uma atividade postada no AVA não foi acessada pelos estudantes pode indicar que: a) o professor não deu instruções adequadas ou b) os conteúdos não estão tendo fácil acesso (problema de visibilidade). Algumas dessas informações estão no próprio AVA, porém de forma dispersa e em base de dados que não ajudam a obter bons resultados de consultas e dinamismo. O professor não tem como saber se o aluno tem acessado o portal de ensino e se estão realmente realizando as atividades virtuais. A pesquisa e implementação do BI nesse ramo, com a implementação de um Data Warehouse(DW), obtendo um ambiente com objetivo único e uma base de dados otimizada para consultas, poderá dar suporte ao entendimento das informações obtidas através do uso do Ambiente Virtual de Aprendizagem para ajudar os professores na avaliação processual, identificando se os métodos aplicados são realmente eficazes num curso EAD e minimizar os problemas já encontrados, como por exemplo, o alto número de evasão de alunos.

O objetivo deste trabalho é elaborar um modelo de Business Intelligence para auxiliar na avaliação processual em cursos de EAD, através de informações úteis disponibilizadas pelo AVA. A criação deste modelo contempla: a construção de um Data Warehouse dimensional, uma arquitetura técnica conceitual e a especificação de um conjunto de visões ou perspectivas de análise de desempenho. A modelagem dimensional será feita com base no levantamento de requisitos de BI, que se constitui de um estudo dos sistemas AVA. Na etapa final serão criadas consultas na base de dados e os resultados serão comparados com os dados gerados nas visões analíticas a fim de validar o modelo proposto neste trabalho.

Este trabalho está estruturado em 4 capítulos. No primeiro capítulo, como foi visto, introduz o problema a ser abordado. O segundo capítulo é feita uma revisão bibliográfica sobre Business Intelligence e Data Warehouse, abordando os temas de acordo com o escopo deste trabalho. No terceiro capítulo é apresentado o planejamento e a metodologia usada neste trabalho, seguido das especificações das arquiteturas técnica e do projeto, a modelagem dimensional e as especificações das visões analíticas, descrevendo o processo de como o modelo foi elaborado. No quarto capítulo é descrito o processo de implementação, validação e resultados. Por fim, no quarto capítulo é feita a conclusão do trabalho relatando os objetivos alcançados, as limitações e as propostas de trabalhos futuros.

2 Fundamentação Teórica

Este capítulo apresenta, de maneira teórica, as técnicas e tecnologias usadas neste trabalho. Tendo como foco maior o Business Intelligence e seus componentes.

2.1 Sistemas de Informação

Desde que os computadores se tornaram equipamentos comerciais, a informação e a tecnologia passaram a caminhar juntas. No princípio, as aplicações eram criadas sem grandes preocupações com a quantidade e duplicidade de dados e processos criados, muito menos com o valor desses dados. Com o avanço tecnológico, o conceito de sistemas de informação surgiu e a qualidade das informações e o seu entendimento passam a estar em evidência. Serão apresentados neste tópico, alguns conceitos que possibilitarão um maior entendimento sobre o poder da informação.

2.1.1 Dados, Informações e Conhecimento

Distinguir o que vem a ser, na prática, dado, informação e conhecimento não é tão simples, visto que eles estão intimamente interligados. Dado é definido por [SETZER \(2001\)](#) como “uma sequência de símbolos quantificados ou quantificáveis.” Que podem ser descritos através de representações formais, estruturais. Portanto, um texto é um dado, bem como, também são dados as fotos, sons, figuras, etc. Os dados são sinais não processados, sem interpretação que podem ser armazenados e processados em um computador.

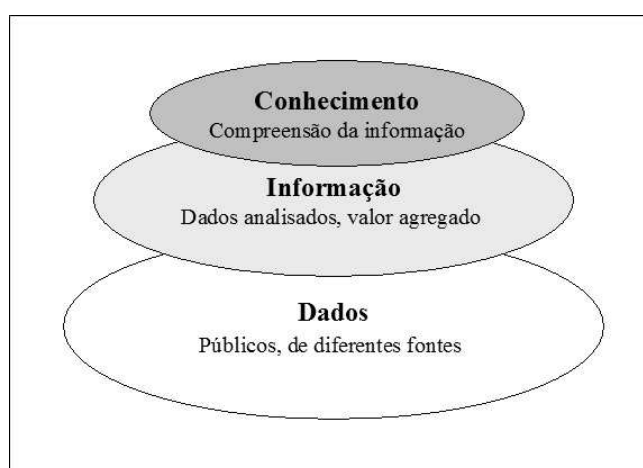
Informação é o significado associado ao dado, uma visão subjetiva sobre o conjunto desses dados. Segundo [SETZER \(2001\)](#), a informação

é uma abstração informal (isto é, não pode ser formalizada através de uma teoria lógica ou matemática), que está na mente de alguém, representando algo significativo para essa pessoa. Note-se que isto não é uma definição, é uma caracterização, porque “algo”, “significativo” e “alguém” não estão bem definidos; assumo aqui um entendimento intuitivo (ingênuo) desses termos. Por exemplo, a frase “Paris é uma cidade fascinante” é um exemplo de informação – desde que seja lida ou ouvida por alguém, desde que “Paris” signifique para essa pessoa a capital da França (supondo-se que o autor da frase queria referir-se a essa cidade) e “fascinante” tenha a qualidade usual e intuitiva associada com essa palavra.

O mesmo conjunto de dados vai gerar informações diferentes para pessoas diferentes, pois dependem das experiências e capacidades individuais. Uma distinção crucial entre informação e dados é que a informação contém semântica e o dado é sintático.

Conhecimento é a compreensão da informação, possibilitando o uso desse conhecimento para algo específico. SETZER (2001) diz que: “Conhecimento é uma abstração interior, pessoal, de alguma coisa que foi experimentada por alguém”. O conhecimento é formado através da informação que está na mente das pessoas, constituído por experiências, interpretações e contextos. Uma informação se torna conhecimento quando uma pessoa consegue ligá-la a outra informação, no contexto específico, por exemplo, se um indivíduo ouviu a seguinte frase: “A temperatura em Toronto é de 2°C” e existe um conhecimento prévio desse contexto, logo serão geradas outras informações como “fará muito frio em Toronto hoje”, “é necessário se agasalhar em Toronto hoje”, etc.

Figura 1 – Dados, Informação e Conhecimento



Fonte: Elaborada pelo Autor.

Portanto, de forma hierárquica, os dados são a base para a informação, que é pré-requisito para o conhecimento, mantendo uma relação de interdependência, como mostrado representado na Figura 1.

Exemplificando esses conceitos discutidos, podemos imaginar um cenário onde um professor observa que um número pequeno de alunos está acessando determinado fórum, diante disso o professor realiza atividades para incentivar os alunos a acessar esse fórum. No primeiro momento o professor observa fatos quantificáveis, o número de alunos, ou seja, os dados. No segundo momento o professor percebe que a quantidade de aluno está pequena, agrupando esses dados, dando-lhes significados. E no terceiro e último momento o professor toma a decisão de realizar atividades para incentivar o acesso ao fórum, decisão essa vinda da junção da informação disponível mais a sua experiência com relação a este cenário.

2.2 Educação à Distância (EAD) e o Moodle

A educação tem como objetivo, formar o ser humano e contribuir para a sua emancipação. A educação a distância tem a mesma funcionalidade. Rompendo com a relação espaço/tempo, que caracteriza a escola presencial, a EAD se concretiza por intermédio da comunicação mediada por meio da mídia. Ao contrário da escola convencional, onde a mediação pedagógica é realizada em contato direto com o aluno, na modalidade à distância a mídia torna-se uma necessidade vital para que a comunicação educacional se concretize. Segundo ALVES e NOVA (2003)

O conceito de EAD remeteria a qualquer modalidade de transmissão e/ou construção do conhecimento sem a presença simultânea dos agentes envolvidos.

Para tornar possível o ensino/aprendizagem por meio da mídia, fazendo a ponte entre o clássico ensino presencial e o ensino a distância, em 1999, surgiu o Moodle. Consiste numa página, onde professores disponibilizam recursos para os alunos, e desenvolvem atividades juntamente com eles. O Moodle é a sala de aula virtual. Uma importante plataforma de gestão de ensino/aprendizagem, de código aberto, bastante flexível, de grande valor educativo e de fácil utilização, graças a sua interface simples, facilitando o uso até de pessoas menos experientes.

Usado pela grande maioria dos cursos EAD, o Moodle apresenta funcionalidades bastante participativas, comunicativas e que estreitam a ligação entre os alunos e professores. Dentre essas funcionalidades de avaliação temos como principais:

- Fórum: é uma ferramenta de discussões de temas, um local onde são feitas reflexões a cerca de um determinado conteúdo, podendo incluir arquivos no mesmo.
- Trabalho: ferramenta onde são postados trabalhos, atividades requisitadas pelos professores para avaliação.
- Chat: é uma ferramenta que facilita a comunicação rápida entre alunos e professores, serve para esclarecimento de dúvidas.
- Questionário: ferramenta que permite construir questionários para os usuários do Moodle, podendo exportar os resultados.
- Teste: a ferramenta permite que o professor crie uma base de dados de perguntas e respostas, podendo escolhê-las aleatoriamente, inclusive corrigir respostas automaticamente.

Com o uso do Moodle e de suas funcionalidades é gerado um grande volume de informações. Essas informações são bastante amplas, elas contemplam desde o simples acesso do aluno ao sistema até as respostas dos alunos nos testes publicados. Essas informações

podem auxiliar numa tarefa de grande dificuldade nos cursos EAD, a avaliação da aprendizagem. Porém, identificar quais informações são realmente importantes e dar significado a elas não é uma tarefa simples.

2.3 Business Intelligence

Desde as primeiras escrituras sobre estratégias é notória a preocupação, em determinado contexto, de conhecermos da forma exata nossos pontos fracos e fortes, bem como os dos nossos adversários ou concorrentes. Ou seja, conhecer é adquirir vantagens. Para isso era necessário analisar dados e gerar informações úteis. Antes da automatização, da introdução dos computadores, a informação era tratada de forma avulsa e não estruturada, as decisões tomadas tinham um grau elevado de intuição, por exemplo, através da observação do comportamento da natureza o homem tomava algumas decisões que lhes proporcionavam benefícios, como a observação do comportamento de alguns animais que indicavam a chegada de um aguaceiro(SEZOES, 2006).

A partir da década de 90 começa a revolução dos dados e da informação, com algumas tendências evidenciadas, as empresas começam a ser induzidas a oferecer aos seus dirigentes melhores informações como forma de preparo para um melhor gerenciamento dos seus negócios. No contexto específico da educação, através dos dados gerados diariamente, as instituições buscam analisar desempenhos de alunos e professores em face de sua didática de ensino. Porém esses dados mudam constantemente e crescem exponencialmente tornando-se complexos, de difícil extração e, conseqüentemente, de pouco valor para apoiar a tomada de decisão.

Nas análises convencionais, por exemplo, nas que são realizadas utilizando-se consultas SQL(Structured Query Language) complexas em banco de dados transacionais, geram comumente lentidão nos sistemas e atendem a uma parcela mínima das necessidades do usuário final.

Neste cenário, para permitir decisões melhores e mais rápidas surge o conceito de Business Intelligence(BI) que é um conjunto de aplicações que possibilitam um acesso rápido, interativo e partilhado das informações. Através do BI os gestores descobrem relações, tendências e transformam grandes quantidades de informação em conhecimento útil. Segundo VERCELLIS (2009) BI é

um conjunto de modelos matemáticos e métodos de análise para explorar dados visando a geração de informações e conhecimentos úteis a processos complexos de tomada de decisão(tradução livre).

O termo, Business Intelligence, foi criado pelo grupo Gartner no início dos anos 1990 e inclui arquiteturas, ferramentas, bancos de dados, aplicações e metodologias que serão abordadas posteriormente. O conceito é bem claro: fazer uso das informações já

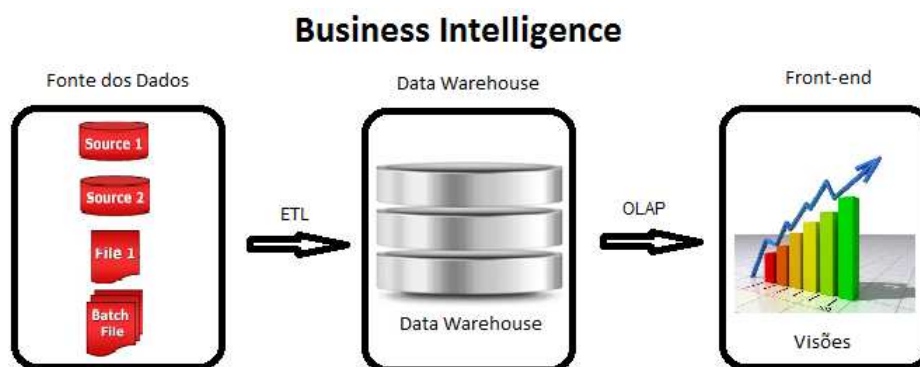
disponíveis da organização para ajudar os gestores a adotar as melhores decisões e de forma rápida.

O BI tem como principal objetivo o acesso a dados confiáveis, disponibilizados em tempo real para um aumento da transparência e compreensão do negócio. Existem algumas ferramentas que juntas compõem o BI, dedicadas a cada etapa específica como por exemplo o processo de extração e padronização dos dados fontes, o data warehouse que irá armazenar os dados, ou ainda a análise dos dados armazenados no data warehouse. Todos esses compõem a arquitetura de Business Intelligence, como será descrito seguir.

2.3.1 Arquitetura e componentes do BI

Um sistema de Business Intelligence, numa perspectiva tecnológica, deve estar inserido na infraestrutura dos Sistemas de Informação das organizações. Porém, deve-se ter sempre em mente que um BI não substitui os sistemas principais, geradores dos dados fontes, pois o BI não subsiste por si próprio, ele está ligado fortemente aos dados gerados pelos sistemas operacionais da organização. Um sistema de BI consiste na implementação de estruturas especiais para armazenar e prover informações para dar suporte a camada de decisão da organização. Essas estruturas são divididas em fonte dos dados, data warehouse e ferramentas de análise de dados, como mostra a Figura 2.

Figura 2 – Arquitetura Business Intelligence



Fonte: Elaborada pelo Autor.

- Fonte de dados: Onde os dados operacionais estão armazenados, podendo esses estar estruturados no banco de dados ou não, por exemplo, dados não estruturados podem estar armazenados em arquivos (.txt), (.csv), etc. Nesta fase o principal objetivo é unificar e padronizar os dados das diferentes fontes. Para isso, são desenvolvidas ferramentas de extração de dados conhecidas como *ETL* (*Extract, Transform and Load*).

- Data Warehouse e Data Marts: Após a extração e transformação dos dados operacionais eles são re-armazenados em estruturas de dados remodeladas, otimizadas para análise, que são conhecidas como Data Warehouse e Data Marts.
- Ferramentas de análise de dados: uma vez consolidado o Data Warehouse, os dados nele inseridos podem ser submetidos a ferramentas específicas que analisam bancos de dados multidimensionais.

Para (KIMBALL, 2002), a arquitetura *Business Intelligence* tem o processo de ETL definido como uma camada exclusiva, dando uma importância maior a essa fase. Kimball define a camada de *Data Staging Area*, onde os dados extraídos para serem tratados vão ser armazenados. Sendo assim a arquitetura proposta por ele se divide em: Fontes dos dados, *Data Staging Area*, *Data Warehouse* e *Data Marts* e a camada de ferramentas que irão analisar os dados modelados dimensionalmente.

2.4 Data Warehouse

O *data warehouse* é um banco de dados que tem como objetivo disponibilizar informações históricas sobre a organização para apoiar a tomada de decisão. Ao contrário dos bancos de dados transacionais operacionais, que tem como foco os dados operacionais gerados das atividades cotidianas da organização, o *data warehouse* tem como foco o conjunto desses dados, organizados de tal forma que deem subsídios para que o gestor possa melhor gerir o seu negócio (KIMBALL, 2002).

Segundo INMON (2002) *data warehouse* é

“uma coleção de dados orientada por assuntos, integrada, variante no tempo, e não volátil, que tem por objetivo dar suporte aos processos de tomada de decisão”.

Tomando como exemplo um sistema de gerenciamento de uma universidade, que seja responsável pela matrícula dos alunos, gerenciamento das matérias a serem cursadas, avaliações a serem feitas pelos alunos, etc. Um banco de dados convencional seria responsável por guardar dados operacionais como os dados dos alunos matriculados, quais as matérias ele está matriculado e as avaliações feitas pelo mesmo, por exemplo. Já um *data warehouse*, para este ambiente, armazenaria informações como a quantidade de alunos matriculados por certos períodos, ou a quantidade de alunos que não fizeram as atividades e que acabaram sendo reprovados na matéria, entre outras informações.

Portanto, já podemos perceber as diferenças evidenciadas entre um *data warehouse* e um banco de dados operacional. Uma dessas diferenças é o usuário. Enquanto o do banco de dados operacional se destina a usuários operacionais da organização, que realizam

atividades cotidianas relativas ao negócio da organização. O *data warehouse* se destina aos gestores, que tem como missão gerir a organização e tomar decisões estratégicas.

É de fundamental importância salientar que os dois bancos de dados devem coexistir. Um *data warehouse* não substitui um banco de dados operacional, não há concorrência entre eles. Visto que, eles são usados para propósitos diferentes, como foi evidenciado anteriormente.

2.4.1 Características do Data Warehouse

Nesta seção serão detalhadas algumas características específicas do *Data Warehouse*, que tornam possível o armazenamento de informações estratégicas passíveis de serem analisadas.

Orientadas por assunto

Ao contrário dos bancos de dados operacionais, que organizam seus dados por aplicação, os DWs são organizados por assunto. A organização por assunto centraliza os dados, com isso, ajuda na recuperação dos dados. O assunto pode ser qualquer tema de interesse, por exemplo, para um sistema de gerenciamento de curso EAD podemos ter como assunto a frequência, atividades, etc. Esta característica está ligada diretamente ao conceito de *Data Mart*, que é um conjunto de dados, modelados dimensionalmente referente a um departamento de uma organização.

Integrada

Os dados inseridos num DW podem vir de diversas fontes, por isso, eles devem ser sincronizados e padronizados. Ou seja, todos os dados devem ser tratados para se obter um padrão único, resultando num dado consistente. Esse processo é denominado como integração de dados. A Figura 3, ilustra a situação onde o DW tem três fontes distintas e cada uma com seu padrão, esses dados são convertidos a um padrão definido e carregados no DW.

Variante no tempo

No DW, a cada ocorrência de mudança, é criada uma nova entrada, para marcar esta mudança. Enquanto que nos bancos de dados transacionais operacionais o dado é atualizado de acordo com a mudança de estado dos objetos, no DW o dado não é atualizado, ele reflete algum momento específico. Dessa forma, construindo um histórico das mudanças sofridas nos dados.

Não volátil

Refere-se ao fato de que no DW só é permitida apenas a carga dos dados e a consulta a estes dados, ambiente chamado de “*load-and-access*”, não existindo volatilidade. Após serem transformados e integrados, os dados são carregados em blocos para o DW.

Figura 3 – Integração dos Dados



Fonte: Elaborada pelo Autor.

Já nos bancos operacionais, os dados são, geralmente, atualizados registro a registro, essa volatilidade vem acompanhada de um esforço considerável para assegurar a integridade e consistência dos dados.

A construção do *data warehouse* com essas características garantem a integração de dados dispersos por diferentes estruturas de bases de dados, e a separação dos dados operacionais dos dados históricos para análise.

2.4.2 Metodologias Data Warehouse

Serão abordadas nesse tópico duas grandes metodologias que estão na literatura clássica de *data warehouse*, a *top-down*, defendida por INMON (2002) e a abordagem *bottom-up*, defendida por KIMBALL (2002) KIMBALL. As duas abordagens defendem a extração de dados diretamente das fontes operacionais, criando uma área temporária de armazenamento, o *Data Staging Area*, para limpeza e tratamento de dados e posterior carga, a partir desta estrutura para o DW (BARBIERI, 2011).

Metodologia top-down Esta abordagem foi proposta por Inmon e da maior ênfase, já de início, na construção do DW, usando ferramentas e técnicas da modelagem relacional para integrar todas as áreas e assuntos da organização. Após a finalização desta etapa que serão criados os *Data Marts*, esses sim de estruturas dimensionais que irão dar suporte as ferramentas de análise de dados do BI.

Metodologia bottom-up Também conhecida como *Star-Schema*, foi proposta por Kimball e se mostra mais simples e com um retorno mais rápido, pois tem como foco, primeiramente, a construção dos *Data Marts* diretamente das fontes operacionais dos dados. Esses *data marts*, que são estruturas dimensionais, são agrupados compondo um *Data Warehouse*. Segundo (GEORGE, 2012) esta abordagem se caracteriza por projetos menores, independentes e incrementais, focando em assuntos específicos e conseguindo

com isso reduzir a complexidade do projeto e dos recursos.

2.5 Modelagem Dimensional

Modelagem dimensional é uma técnica de modelagem de banco de dados que permite uma melhor compreensão de um modelo de dados por nós seres humanos e que possibilita analisar os dados de N dimensões no formato de um “cubo”. O grande diferencial dos *data warehouse* é exatamente o uso da modelagem dimensional. É este modelo que lhe assegura suas vantagens, tais como sua melhor navegabilidade e seu ganho de performance (KIMBALL, 2002).

A modelagem dimensional tem uma estrutura simples, sendo composta por apenas dois tipos de tabelas, as tabelas fato e dimensão, que possuem características extremamente distintas que serão abordadas a seguir.

2.5.1 Tabelas Fato

A tabela fato é considerada como a tabela mais importante do modelo dimensional. Nela são armazenados os medidores numéricos de desempenho da organização, o fato a ser medido. Numa tabela de fatos, cada linha representa uma medição, sendo que todas as linhas devem ter a mesma granularidade. Esta granularidade de cada linha é referenciada como o grão da tabela fato (KIMBALL, 2002).

Vale salientar que, as ferramentas de análise de dados, dificilmente buscam uma única linha por vez de uma tabela fato, normalmente são buscadas diversas linhas. Por isso, os fatos devem da suporte a operações de agregação sobre eles, como a soma, média etc. A tabela fato possui uma lista de referências para as tabelas de dimensão, a interseção entre as duas tabelas define-se a granularidade e o escopo de uma medição (KIMBALL, 2002).

2.5.2 Tabelas Dimensão

As tabelas dimensionais contêm, na maioria das vezes, muito menos registros do que as tabelas de fato. O seu tamanho em linhas varia habitualmente entre dezenas, centenas podendo chegar a milhares, já nas tabelas fato poderão conter milhões de registros.

As tabelas de dimensão contêm descritivos textuais sobre o negócio, que detém uma chave de substituição à chave primária das fontes de dados que é chamada “chave substituta” ou “surrogate key” (KIMBALL, 2002).

Na tabela de dimensão, os atributos são de extrema importância, visto que, eles funcionam como restrições sobre consultas, tendo um papel de agrupamento e rótulo de relatórios. Por exemplo, suponha que um coordenador de um curso deseje consultar o

número de alunos de um determinado curso que tenham realizado certa matéria e tenham sido reprovados. Para que essa consulta seja possível os atributos “curso” e “matéria” devem estar disponíveis em tabelas de dimensão.

Esse modelo simples, de apenas dois tipos de tabelas é desejável, pois faz com que tenhamos um ambiente otimizado para consultas. As tabelas fatos são sempre normalizadas, porém as tabelas dimensão estão mais sujeitas a serem desnormalizadas. Existem duas linhas sobre a normalização ou não de tabelas dimensão, que serão abordadas a seguir.

O **Modelo Estrela** é talvez o modelo mais utilizado na concepção de um *data warehouse*, este modelo se caracteriza pela não recomendação da normalização das tabelas dimensão. Tem como principal característica o fato de todas as suas dimensões estarem vinculadas somente a tabelas fato. Esta característica decorre exatamente da ausência de normalização.

O **Modelo Floco-de-Neve ou Snowflake**, se caracteriza pela normalização das tabelas dimensão. A principal intenção deste modelo com a normalização é a não redundância de dados e um melhor aproveitamento do espaço em disco. Porém, a normalização vai impactar diretamente no desempenho das consultas, problema existente pela necessidade das operações de junção demandadas por esse modelo.

2.6 *Extract, Transform and Load – ETL*

Os dados fontes que serão carregados no *data warehouse* provirão das mais variadas fontes, cada uma com seus padrões estabelecidos. Antes de carregar esses dados é necessário submetê-los a um processo de padronização e integração até que estes dados estejam aptos para serem carregados no DW, a esse processo dá-se o nome de Extração, Transformação e Carga (ETC) ou *Extract, Transform and Load* (ETL). Neste momento será analisado com mais detalhes cada etapa do processo de ETL.

- **Extração** A fase de extração tem como objetivo coletar os dados de diversas fontes que serão necessários para o DW. Estas fontes, também chamadas de *data sources*, podem ser planilhas, arquivos de textos, porém normalmente são bancos de dados operacionais da organização.
- **Transformação** Como os dados são obtidos de diversas fontes, não há como garantir a qualidade dos mesmos nem uma padronização. É nessa fase que os dados são “lapidados”, ou seja, ocorre uma limpeza e transformação dos dados já extraídos visando tratar inconsistências, imprecisões e padronizar esses dados. Nessa fase, esses dados são armazenados numa região temporária de uso auxiliar para a qual dá-se o nome de *Data Staging Area* (DSA), uma espécie de repositório temporário. Um

exemplo de dado não padronizado é quando encontramos mais de uma formatação de data, como dd/mm/aaaa, dd-mm-aa, etc.

- **Carga** Uma vez que os dados já estão “lapidados”, estes podem ser carregados para o DW, caracterizando a última etapa do processo de ETL. As cargas posteriores a primeira levam consigo alguns cuidados, devendo estas ser submetidas a um processo de sincronia e não volatilidade do conteúdo do DW para com as alterações nos data sources. A carga pode assumir um alto grau de complexidade se considerarmos a necessidade de se garantir a integridade dos dados, isto é, conferir integridade das chaves estrangeiras com chaves primárias. O processo de carga pode ser feito de modo incremental ou total. Na carga incremental, apenas os dados modificados serão carregados, normalmente ocorrem nas tabelas fato, devido ao seu volume de dados muito maior. Já na carga total, todos os dados da tabela de destino são excluídos e sobrescritos, usualmente nas tabelas dimensão.

O processo de ETL é considerado por muitos como uma das fases mais complexas, apesar de já existirem no mercado diversas ferramentas que auxiliam este processo. É um processo de extrema importância, pois é ele que vai garantir o atendimento da característica de não volatilidade do *data warehouse*.

2.7 On-Line Analytic Processing - OLAP

OLAP (On-Line Analytic Processing) ou processamento analítico on-line é uma tecnologia que permite os gestores analisar e visualizar os dados estruturados em cubos de forma rápida, consistente e principalmente interativa. Segundo (VERCELLIS, 2009), OLAP consiste em todo o conjunto de ferramentas voltadas para realizar análises de Business Intelligence e dar suporte a camada de decisão.

As ferramentas OLAP são geralmente desenvolvidas para trabalhar com banco de dados desnormalizados, sendo capazes de navegar pelos dados de um *data warehouse*, possuindo uma estrutura ideal tanto para pesquisas como para a apresentação de informações.

Os sistemas transacionais ou OLTP, como o nome sugere, são sistemas que se baseiam em transações, tem como propósito realizar as atividades relacionadas ao operacional da organização. Já o OLAP tem como propósito dar suporte ao BI e a tomada de decisão.

Serão abordados alguns termos referentes a operações OLAP que estão relacionados a análise multidimensional dinâmica dos dados, apoiando o usuário final. Dentre as diversas operações e características do OLAP, serão abordados as principais e as que serão referenciadas neste trabalho.

Drill Across

O *drill across* acontece quando um usuário “salta” um nível intermediário dentro de uma mesma dimensão. Por exemplo, na dimensão tempo tem o ano, trimestre, semestre, mês e dia. O usuário estará executando um *drill across* quando numa análise da frequência dos alunos por ano, ele passa a analisar essa frequência mensalmente.

Drill Up ou Roll Up

É quando, numa operação de agregação de dados, o usuário sai de um nível mais específico para um mais genérico dentro de uma hierarquia. Segundo BARBIERI (2011), *Drill Up* está diretamente relacionado com o fato de sair de um nível mais baixo da hierarquia para um mais alto, conforme ilustrado na Figura 4. Voltando ao exemplo da frequência no tempo, o *drill up* acontece quando o gestor, que está analisando a frequência diária dos alunos, passa analisar a frequência mensal.

Figura 4 – Drill Up



Fonte: Adaptada de (PASCHOAL, 2012).

Drill Down ou Roll Down

Se trata do inverso do *drill up*, Segundo VERCELLIS (2009), consiste na navegação a partir de dados mais genéricos para dados mais específicos. Como mostra na Figura 5, seria sair de um nível mais alto da hierarquia para um mais baixo.

Figura 5 – Drill Down



Fonte: Adaptada de (PASCHOAL, 2012).

Slicing e Dicing

Slicing e *dicing* são operações OLAP que combinam diferentes perspectivas. O *slicing* permite fixar um valor de uma dimensão para análise. Já o *dicing*, obtém o sub-cubo no espaço do modelo dimensional através da combinação de duas ou mais operações de *slicing*.

3 Modelo de BI para Análise de Desempenho dos Alunos em Cursos EAD

Neste capítulo será especificado o modelo de *Business Intelligence* para análise de desempenho dos alunos em cursos de educação à distância, modelo ao qual esse trabalho se propõe. O modelo aqui proposto tem como componentes: um modelo de dados dimensional, uma arquitetura técnica de especificação e perspectiva de análise. Para sua elaboração foi feito um levantamento de requisitos que consistiu na instalação, configuração e uso do Moodle para análise da realidade dos dados.

3.1 Desenvolvimento do Modelo

Esta seção está subdividida em duas subseções, a primeira que descreve o planejamento e segunda a metodologia.

3.1.1 Planejamento do Modelo

Como primeiro passo para o desenvolvimento do modelo foi definido aqui um planejamento para construção do projeto com o intuito de definir as atividades envolvidas em sua construção.

O planejamento será apresentado sob a forma de uma estrutura analítica de projeto (EAP). De acordo com (PMI, 2008) o EAP representa uma

“ decomposição hierárquica orientadas a entregas do trabalho a ser executada pela equipe para atingir os objetivos do projeto e criar as entregas requisitadas, sendo que cada nível descendente da EAP representa uma definição gradualmente mais detalhada da definição do trabalho do projeto.”

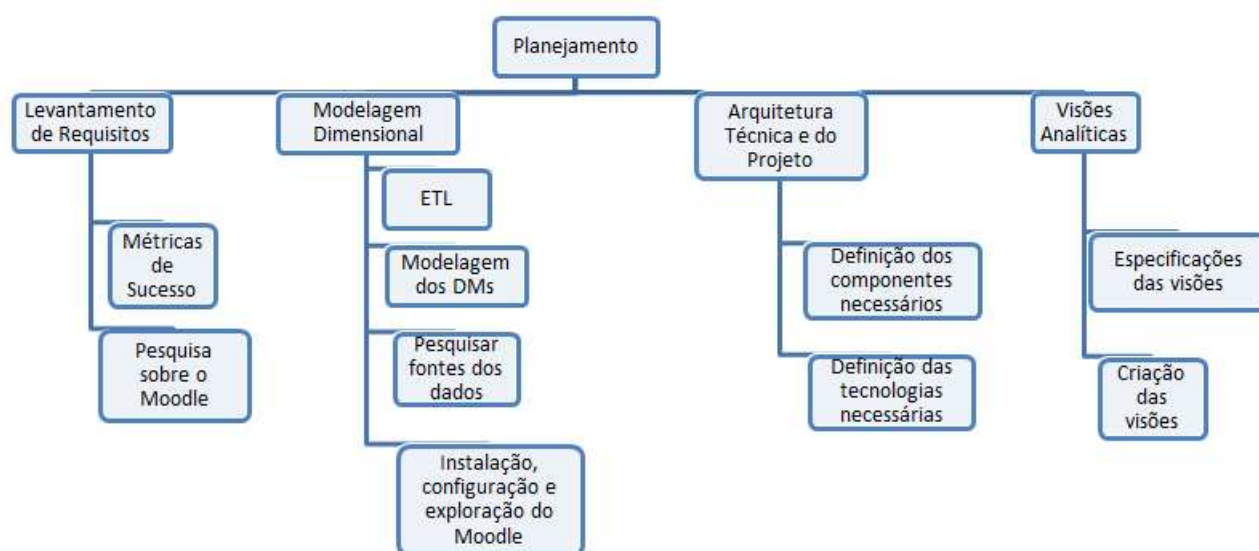
Por meio de uma espécie de organograma, a EAP representa o que deverá ser entregue pelo projeto. Por se tratar de uma pesquisa acadêmica, optou-se por definir formalmente o projeto para que fique claro o objetivo e o produto final deste trabalho. Portanto, esse trabalho consiste na elaboração de um modelo de Business Intelligence para análise de desempenho dos alunos de um curso EAD que utilizam o Moodle como portal de gerenciamento acadêmico do curso.

Dos componentes deste modelo, a arquitetura técnica é a definição dos componentes necessários, que irão compor a arquitetura do BI. O modelo dimensional de dados é o que constituirá o DW e apoiará as ferramentas de análise, e por último, a camada de visão que assumirá o papel de apresentar as informações para análise por parte dos usuários.

De acordo com o foco do trabalho definiu-se que os objetos deste trabalho são: Pesquisa sobre o Moodle para levantamento de requisitos; ETL; *Data Warehouse e Data Marts*, Camada de visões analíticas.

Definidos os objetos deste trabalho, eles foram relacionados às atividades necessárias para sua construção. Na Figura 6 estão representadas graficamente essas atividades num diagrama de Estrutura Analítica de Projeto (EAP).

Figura 6 – EAP



Fonte: Elaborada pelo Autor.

3.1.2 Metodologia do Modelo

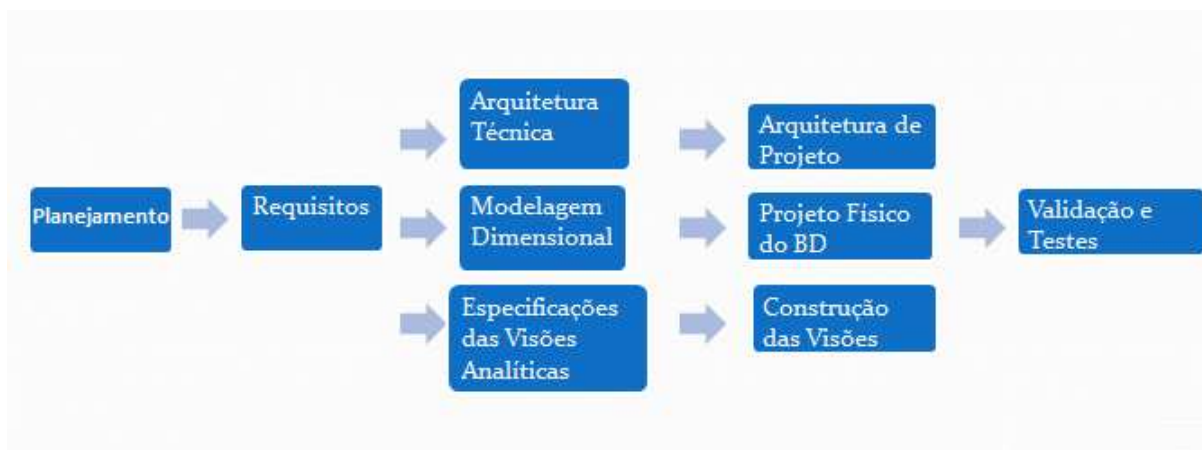
Com o intuito de esclarecer como as atividades do item anterior irão ser conduzidas, será descrito agora a metodologia adotada para o desenvolvimento do modelo.

A metodologia adotada por este trabalho foi proposta por [KIMBALL \(2002\)](#) e ficou conhecida como metodologia *bottom-up*. Ela se caracteriza como uma metodologia de baixa complexidade e retorno mais rápido, visto que possui como foco a construção dos *Data Marts* diretamente das fontes operacionais dos dados. Sendo ideal para projetos menores, independentes e incrementais.

Tomando como base a imagem “mapa rodoviário”, usada por Kimball, que define uma sequência de macro atividades a serem realizadas, foi criada a Figura 7, adaptada, com as atividades que são contempladas por este projeto.

A primeira etapa refere-se ao planejamento, etapa esta já realizada no item anterior. Na etapa de Requisitos, foram extraídas informações do Moodle através da instalação,

Figura 7 – Metodologia



Fonte: Adaptada de (KIMBALL, 2002).

configuração, e do uso da tecnologia para entender o funcionamento do mesmo e identificar as possíveis informações que dão suporte ao usuário na avaliação processual do aluno.

Na etapa seguinte, na arquitetura técnica foram especificados os componentes necessários ao modelo, os elementos que compuseram a arquitetura de BI e suas interações. Na modelagem dimensional, foram modelados os *data marts* definidos a partir dos requisitos obtidos. Na especificação das visões analíticas foram definidas algumas perspectivas de visualização e análise de dados como ponto de partida para que os usuários obtivessem informações importantes para uma satisfatória avaliação processual dos alunos.

Na etapa posterior, a arquitetura de projeto foi desenvolvida, que consiste na definição das tecnologias usadas no projeto, e logo em seguida no projeto físico do BD foram instaladas as tecnologias necessárias para posteriormente serem criadas as tabelas e suas relações, bem como as tecnologias usadas para a construção das visões específicas em ferramentas OLAP.

Na última etapa, foi finalmente realizada a carga dos dados no DW, tornando possível a realização de consultas de teste, consultas essas feitas primeiramente diretas no banco de dados para posterior comparação com dados mostrados nas visões analíticas validando o modelo. Na próxima seção, optou-se por descrever a arquitetura técnica e a arquitetura do projeto juntas para um melhor entendimento, pois apresenta os elementos que irão compor a arquitetura do modelo ao mesmo tempo em que são especificados esses elementos.

3.2 Arquitetura Técnica e do Projeto

Na arquitetura técnica são especificados os componentes necessários ao modelo, os elementos que compuseram a arquitetura de BI e suas interações. Já na arquitetura do projeto são detalhados os componentes usados no modelo, quais tecnologias de fato foram usadas. Na elaboração desta arquitetura foi analisada cada camada contemplada neste trabalho: fontes de dados, ETL, *data warehouse* e *data marts*, e as aplicações de tratamento de dados.

3.2.1 Camada de fontes de dados

Como visto na seção 2.3.2, a fonte de dados é onde os dados operacionais estão armazenados. A fonte de dados deste trabalho é o banco de dados que armazena as informações do MOODLE, que utiliza a tecnologia PostgreSQL que é um sistema gerenciador de banco de dados(SGBD) objeto relacional avançado de código aberto.

O Moodle é composto por módulos e o banco de dados reflete muito bem isso, as tabelas foram nomeadas com um prefixo padrão que é a concatenação da sigla “mdl” mais o nome do módulo. Por exemplo, a tabela principal que armazena algumas informações sobre os fóruns é a *mdl-forum* e as secundárias são: *mdl-forum-discussions*, *mdl-forum-posts* etc. O Banco do Moodle conta com 199 tabelas, um banco razoavelmente grande e de difícil entendimento.

3.2.2 Extract, Transform and Load - ETL

Nesta fase, foi feita a extração dos dados a partir da camada de fontes de dados, representado pelo banco de dados que armazena as informações do Moodle.

Com o uso da IDE NetBeans IDE 7.4, foi desenvolvido um algoritmo, em Java, para extrair essas informações, nesse algoritmo foi feita a conexão com o PostgreSQL e foram executados comandos SQL para buscar essas informações, uma vez de posse desses dados eles foram tratados para se adequar ao DW.

3.2.3 Data Warehouse e Data Marts

Para a criação do *Data Warehouse* foi usado o banco de dados MySQL, que é um SGBD de excelente desempenho e estabilidade. Com grande portabilidade e compatibilidade com diversas linguagens de programação, o MySQL, tecnologia diferente da usada pela fonte de dados, foi escolhido também para deixar mais evidente o fato de o *Data Warehouse* ser um banco de dados a parte do operacional. Os *Data Marts* serão detalhados mais a frente na seção 3.3.

3.2.4 Especificações das Visões Analíticas

Nesta fase a ferramenta OLAP usada foi a Pentaho, uma plataforma de código aberto, desenvolvida em Java, que proporciona ferramentas de BI que auxiliam na gestão da informação. O Pentaho (2005) conta com uma versão open source, o Pentaho Community Edition que tem um ambiente integrado e de fácil manuseio.

O pentaho possui um servidor OLAP, o Mondrian (2005), que tem a capacidade de manipular e analisar um grande volume de dados sob várias perspectivas. Ele interpreta arquivos XML que mapeiam um banco de dados multidimensional, que são os cubos. Para cada *Data Mart* foi criado um esquema de cubo, a Figura 8 mostra o mapeamento do cubo Frequência, onde foram definidas as dimensões com suas hierarquias e métricas.

Figura 8 – Mapeamento dos cubos.

```
<?xml version="1.0"?>
<Schema name="SchemaBIEAD">
+ <Cube name="Quiz" enabled="true" cache="true">
- <Cube name="Frequencia" enabled="true" cache="true">
  <Table name="frequencia"> </Table>
  - <Dimension name="Tempo" foreignKey="ID_TEMPO" type="StandardDimension">
    - <Hierarchy name="Tempos" primaryKey="ID" allMemberName="Todos" hasAll="true">
      <Table name="tempo"> </Table>
      <Level name="Ano" type="Numeric" hideMemberIf="Never" levelType="Regular" uniqueMembers="false" column="ANO"> </Level>
      <Level name="Semestre" type="Numeric" hideMemberIf="Never" levelType="Regular" uniqueMembers="false" column="SEMESTRE"> </Level>
      <Level name="Trimestre" type="Numeric" hideMemberIf="Never" levelType="Regular" uniqueMembers="false" column="TRIMESTRE"> </Level>
      <Level name="Mes" type="Numeric" hideMemberIf="Never" levelType="Regular" uniqueMembers="false" column="MES"> </Level>
      <Level name="Semana" type="Numeric" hideMemberIf="Never" levelType="Regular" uniqueMembers="false" column="SEMANA"> </Level>
      <Level name="Dia" type="Numeric" hideMemberIf="Never" levelType="Regular" uniqueMembers="false" column="DIA"> </Level>
    </Hierarchy>
  </Dimension>
  - <Dimension name="Aluno" foreignKey="CPF_ALUNO" type="StandardDimension">
    - <Hierarchy name="Aluno" primaryKey="CPF" allMemberName="Todos os Alunos" hasAll="true">
      <Table name="aluno"> </Table>
      <Level name="Aluno" type="String" hideMemberIf="Never" levelType="Regular" uniqueMembers="false" column="NOME"> </Level>
    </Hierarchy>
  </Dimension>
  - <Dimension name="Curso_Disc" foreignKey="ID_CURSO" type="StandardDimension">
    - <Hierarchy name="Curso_Disc" primaryKey="ID" allMemberName="Todos oos Cursos" hasAll="true">
      <Table name="curso"> </Table>
      <Level name="Curso" type="String" hideMemberIf="Never" levelType="Regular" uniqueMembers="false" column="CURSO"> </Level>
    </Hierarchy>
  </Dimension>
  <Measure name="Dias_Sem_Acesso" column="DIAS_SEM_ACESSO" visible="true" aggregator="sum" datatype="Numeric"> </Measure>
</Cube>
</Schema>
```

Partindo das informações mais importantes contidas em cada cubo de dados, definimos as visões analíticas deste trabalho, um ponto de partida para os usuários, que serão apresentadas em forma de tabelas e gráficos. Foram criadas cinco visões sendo três relacionadas ao cubo do QUIZ e duas relacionadas ao cubo de FREQUÊNCIA, que serão detalhadas na seção 3.4.

3.3 Modelagem Dimensional

Seguindo a metodologia *bottom-up*, o *data warehouse* é formado por um conjunto de *data marts*, como foi visto anteriormente. Ou seja, a modelagem do *data warehouse*

é o conjunto de modelagem dos seus *data marts*, caracterizando o *bottom-up* como uma metodologia simples, incremental e integrada (KIMBALL, 2002).

A modelagem dimensional, neste trabalho, foi desenvolvida através do levantamento de métricas extraídas do Moodle, métricas que servirão de apoio aos usuários do DW para auxiliar na avaliação processual do aluno. Segundo Kimball, para a construção do modelo dimensional é necessário seguir 4 etapas:

- **Selecionar o processo de negócio**
- **Declarar o grão do processo de negócio**
- **Escolher as dimensões**
- **Identificar os fatos numéricos da tabela de fatos**

Como processo de negócio foram escolhidas as atividades *Quiz* e análise da Frequência. Poderia ser escolhida qualquer outra atividade, essa escolha foi aleatória, pois não é foco do trabalho identificar qual atividade é mais vantajosa para uma avaliação processual, o foco do trabalho é mostrar que é possível e viável a utilização da tecnologia Business Intelligence para auxiliar os professores na avaliação processual de alunos de um curso EAD. Sendo assim, nas próximas seções iniciaremos a modelagem dos *datas marts*, começando pelo DM de Frequência.

3.3.1 Data Mart Frequência

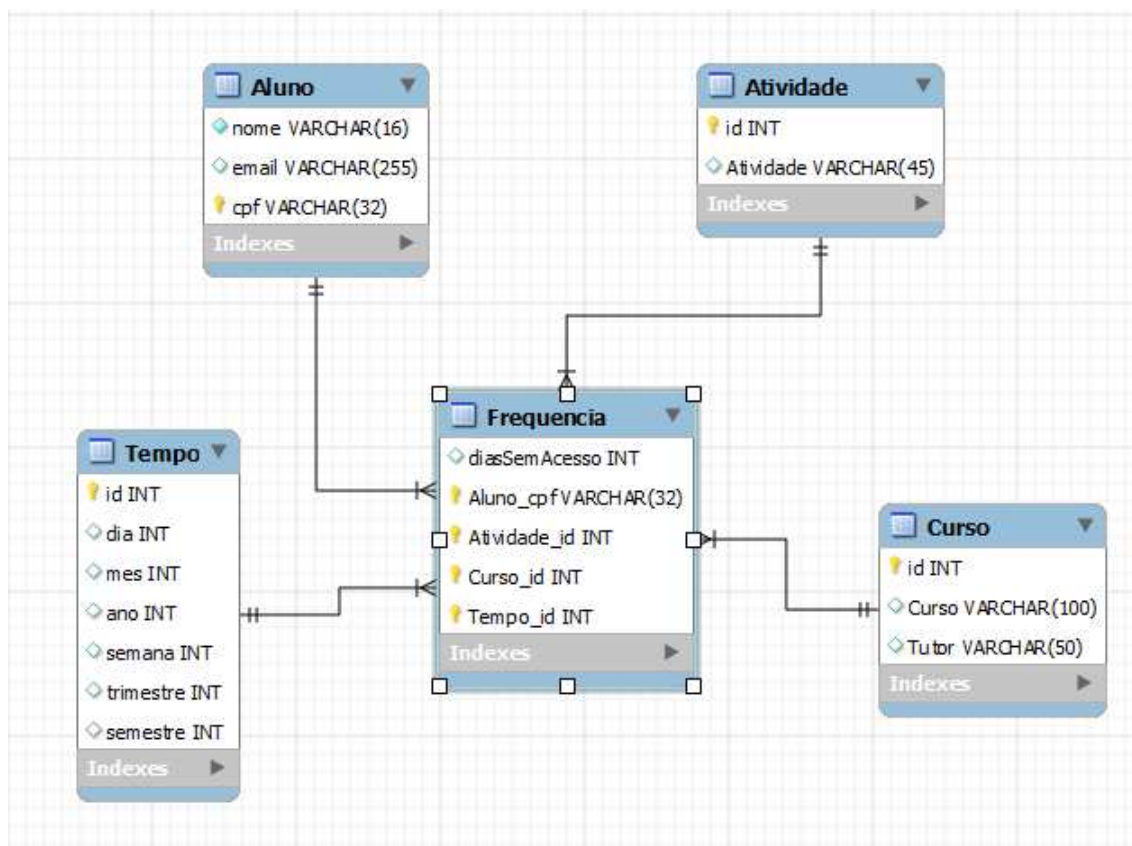
O *data mart* frequência detém a informação da frequência com que o aluno está acessando as atividades postadas no Moodle. Assim como o professor de um curso presencial tem como fazer uma avaliação processual identificando os alunos que estão com dificuldade na matéria ou se ausentando da mesma, o uso deste cubo dar suporte ao acompanhamento do acesso dos alunos ao moodle de forma gradual.

A granularidade é o nível de detalhes de dados que ficarão disponíveis no *data mart*, é representada pela linha da tabela de fatos. O grão definido para este DM foi declarado como a ação de acessar uma atividade no Moodle. Através dessa informação o professor pode saber se os alunos estão acompanhando as atividades, se estão deixando de fazê-las podendo investigar quais são as dificuldades encontradas.

A modelagem do DM Frequência está representada na Figura 9, contemplando as especificações de tipo e seus tamanhos, informações estas que já caracterizam uma implementação parcial da mesma.

A dimensão Aluno contém os campos nome e cpf, sendo o cpf o campo que representa a chave primária. Na dimensão tempo, há referência à data de acesso do aluno

Figura 9 – Data Mart de Frequência



Fonte: Elaborada pelo Autor.

as atividades, a típica tabela dimensão que tem como característica a restrição sobre a consulta. E nas tabelas Curso e Disciplina, temos apenas os seus descritores.

E finalmente, na tabela fato Frequência teremos reunidas as referências para o curso, o aluno, o dia, a atividade e a o campo “diasSemAcesso”. O último diz respeito ao número de dias sem acesso ao moodle.

Definido o *data mart* frequência, agora definiremos o *data mart Quiz*.

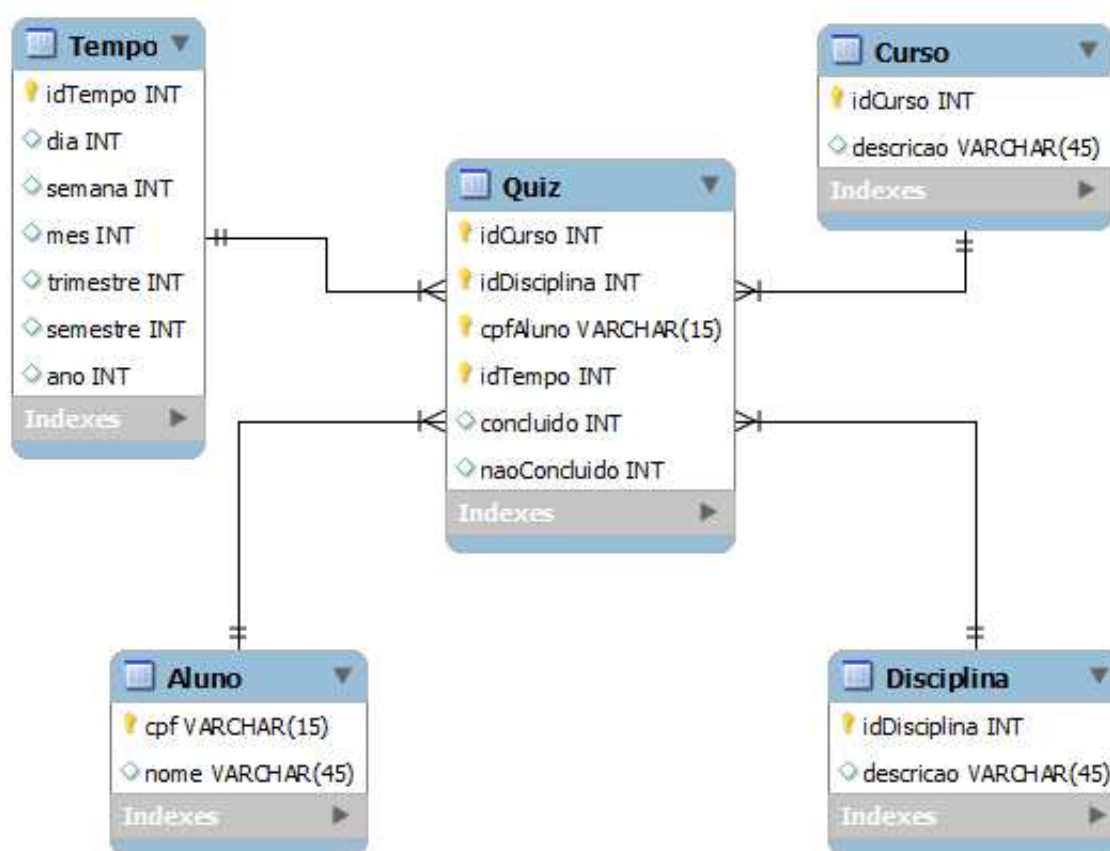
3.3.2 Data Mart Quiz

Os *quiz* são questionários que tem como finalidade avaliar o conhecimento dos alunos em determinado assunto. Sendo assim, o grão definido para o DM *Quiz* foi à verificação diária do status de andamento do *quiz*, podendo ser concluído ou não concluído. Com essa informação, o professor terá um feedback do andamento do *quiz*, se os alunos estão realmente realizando aquela atividade.

A modelagem do DM *Quiz* está representada na Figura 10, onde foi adotado um diagrama de entidade-relacionamento visando descrever o modelo de dados, contemplando as especificações de tipo e seus tamanhos, sendo algumas dimensões já citadas anterior-

mente.

Figura 10 – Data Mart de Quiz



Fonte: Elaborada pelo Autor.

Neste *Data Mart*, a única dimensão não conhecida é disciplina, que conta com um campo descrição, que vai servir para restringir as consultas a determinada disciplina. Na tabela fato, temos a referência para as dimensões Aluno, Tempo, Curso e Disciplina. Além dessas referências, a tabela fato contém duas métricas que condizem ao status do *quiz* que está sendo realizado pelo aluno em determinado momento, sendo as métricas: concluído e não concluído.

Explicado o *Data Mart Quiz*, que tem como grão à verificação diária do status de andamento do *quiz* realizado por um determinado aluno, de um curso, de certa disciplina, em um determinado tempo. Na seção seguinte vamos Especificar das Visões Analíticas.

3.4 Especificação das Visões Analíticas

Esta fase contempla apenas uma ferramenta OLAP, visto que se trata de um projeto pequeno e a intenção é de gerar um conjunto inicial mínimo de visões analíticas.

Como citado na seção 3.2.4, a ferramenta usada foi o Pentaho, escolhida por ter amplo uso, boa documentação e contar com uma distribuição open source. O uso de visões analíticas facilita na usabilidade de uma ferramenta OLAP, visto que são criadas visões específicas com informações diretas que são de melhor e rápido entendimento. É importante salientar que as visões não substituem a livre exploração da ferramenta OLAP, que dá maiores possibilidades de combinações de informações. Partindo das informações mais importantes contidas em cada cubo de dados, definimos as visões analíticas deste trabalho que serão um ponto de partida para os usuários, serão apresentadas em forma de tabelas e gráficos. Foram criadas cinco visões sendo três relacionadas ao cubo do *QUIZ* e duas relacionadas ao cubo de FREQUÊNCIA.

3.4.1 Variação da quantidade de alunos que concluíram o *quiz* por mês

Esta visão tem como objetivo fornecer ao usuário uma análise sobre a variação da quantidade de alunos que concluíram o *quiz* em determinado mês e ano escolhido. Se trata de uma informação genérica, que dá um apanhado geral de todos os alunos da instituição que iniciaram o uso do *quiz* naquele intervalo de tempo escolhido. A visão foi representada através da tabela na Figura 11 a seguir.

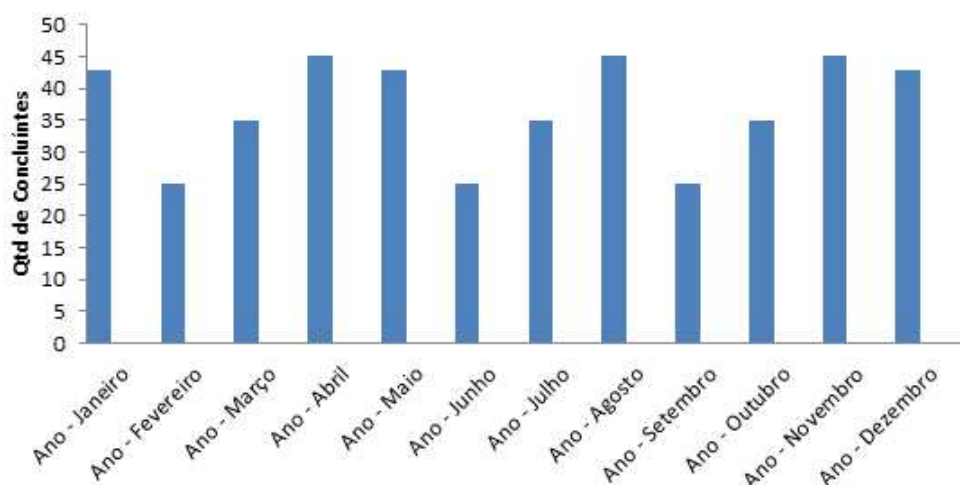
Figura 11 – Variação da quantidade de alunos que concluíram o quiz por mês

Ano	Mes	Quiz Concluido

Fonte: Elaborada pelo Autor.

Foi especificada também para essa visão a criação de um gráfico onde cada barra vertical representa o ano e o mês, e a sua altura representa a quantidade de concluintes da atividade do *quiz*. Segue a representação no gráfico da Figura 12.

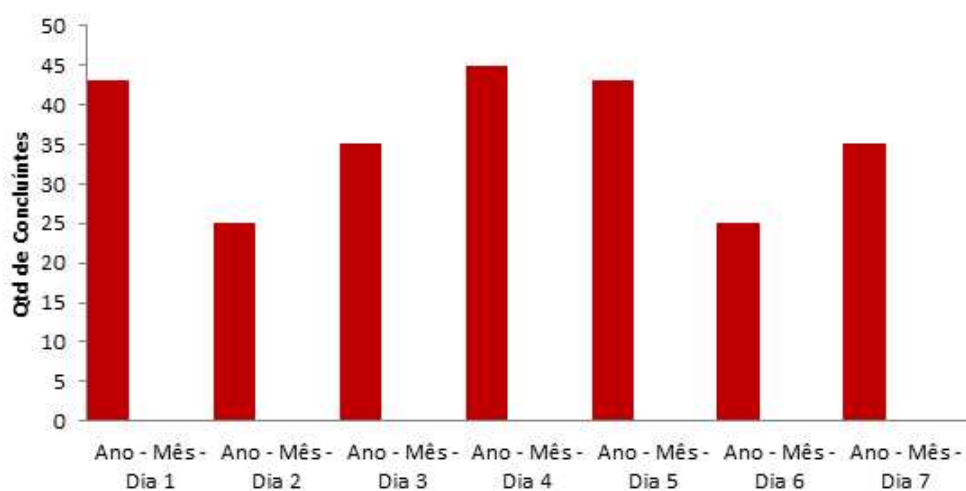
Figura 12 – Variação da quantidade de concluintes.



Fonte: Elaborada pelo Autor.

Como visto na seção 2.7 a ferramenta OLAP tem diversas operações e características, uma dessas operações é a Drill Down que consiste na “navegação” de um dado mais genérico para um mais específico, para conferir essa característica basta dar um duplo-click em uma das barras do gráfico para que o mesmo mude e passe a apresentar os dados apenas daquele mês organizado pela semana, ao dar um duplo-click outra vez será aprofundado em mais um nível, mostrando agora apenas daquela semana organizado pelos dias como mostra na Figura 13.

Figura 13 – Variação da quantidade de concluintes da primeira semana do mês.



Fonte: Elaborada pelo Autor.

3.4.2 Percentual de alunos que concluíram o *quiz* no mês

Essa visão tem como objetivo apresentar o Percentual de alunos que concluíram o *quiz* em determinado mês. Se assemelha bastante com a visão anterior, porém trás um dado complementar que dá um maior entendimento, pois mostra qual a quantidade de alunos estão realmente conseguindo finalizar a atividade com relação a totalidade da turma. Representada na tabela da Figura 14.

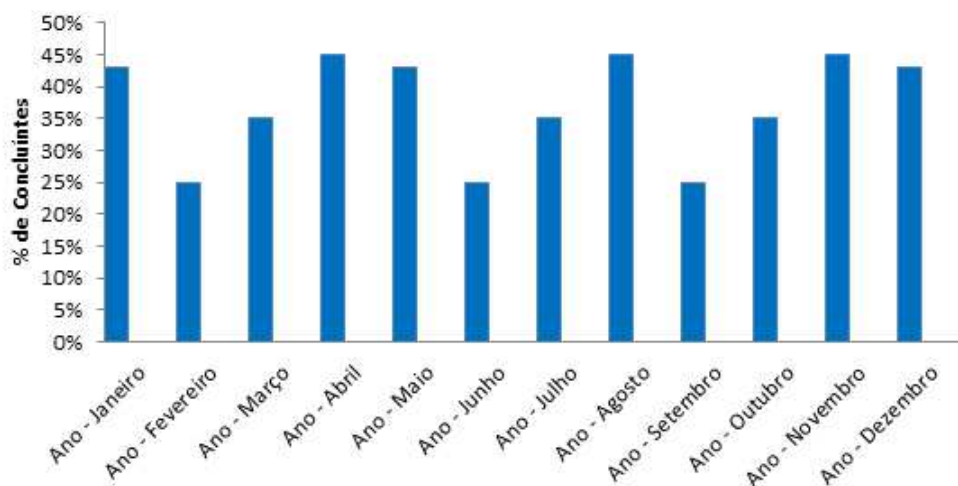
Figura 14 – Percentual da quantidade de alunos que concluíram o quiz por mês

Ano	Mes	Percentual %

Fonte: Elaborada pelo Autor.

Para essa visão, também foi confeccionado um gráfico em barras, representado na Figura 15, onde inicialmente cada barra representa o ano e o mês, e a sua altura a porcentagem da quantidade de alunos que concluíram o quiz naquele mês, dando suporte também as operações de uma ferramenta OLAP, como o Drill up e o Drill Down.

Figura 15 – Percentual da quantidade de concluintes.



Fonte: Elaborada pelo Autor.

3.4.3 Quantidade de alunos que concluíram o *quiz* por curso

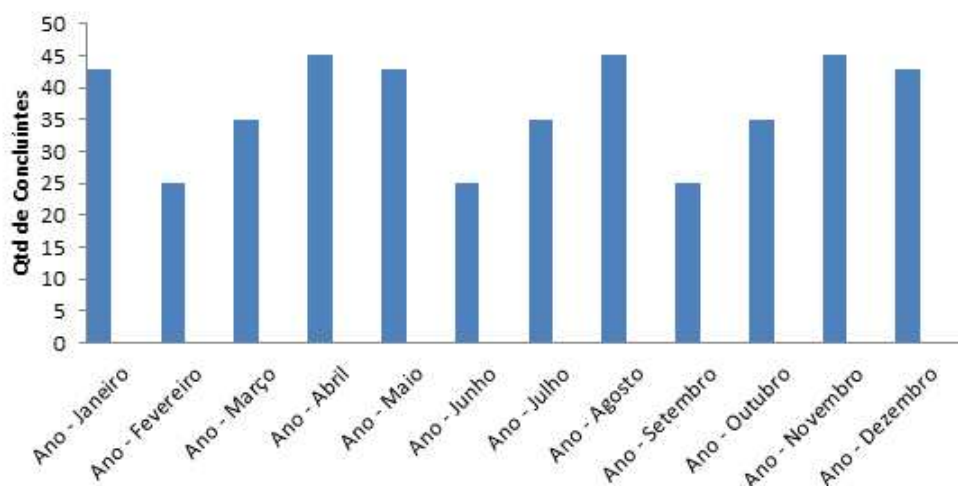
Nessa visão foi inserida uma nova dimensão. O usuário terá como ponto de partida a quantidade de alunos que concluíram o *quiz* de cada curso num determinado ano, como pode ser visto na tabelada Figura 16 e no gráfico da Figura 17. Dessa forma o usuário pode identificar em qual curso em específico pode estar ocorrendo o não comprimento das atividades de *quiz* e tomar providências com um maior direcionamento.

Figura 16 – Quantidade de alunos que concluíram o quiz por curso

Ano	Curso	Quiz Concluido

Fonte: Elaborada pelo Autor.

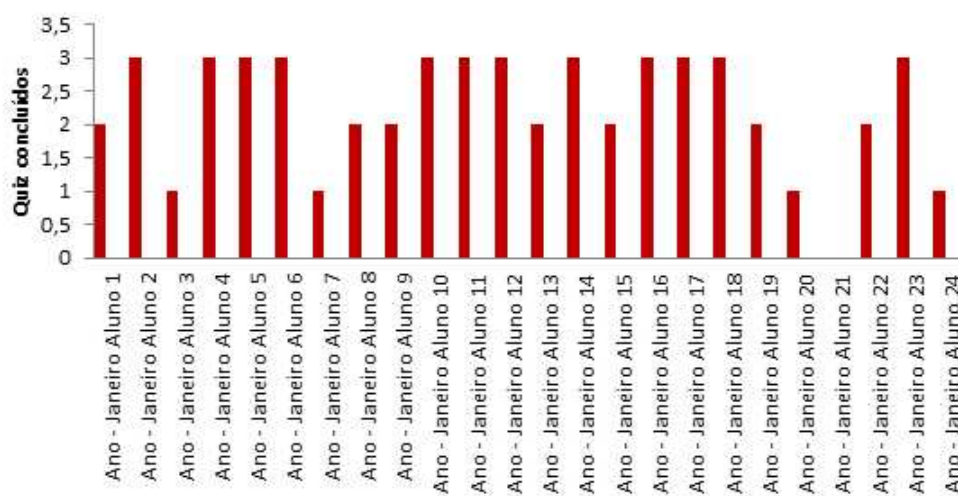
Figura 17 – Quantidade de alunos que concluíram o quiz por curso no ano.



Fonte: Elaborada pelo Autor.

Ao analisar as visões o usuário pode aumentar ou diminuir a granularidade à medida que ele vai colocando ou retirando dimensões no gráfico, por exemplo, o usuário pode aumentar a granularidade adicionando a dimensão aluno, dessa forma, uma quantidade maior de informações serão apresentadas no gráfico, como se pode observar no gráfico da Figura 18.

Figura 18 – Quantidade de quiz que foram concluídos de cada aluno no ano e mês.



Fonte: Elaborada pelo Autor.

3.4.4 Quantidade de dias que os alunos ficaram sem acessar o portal por Tempo

Esta visão é referente ao cubo de Frequência, ela mostra a quantidade de dias que os alunos passaram sem acessar o moodle, como mostra a tabela da Figura 19. No gráfico da Figura 20, a altura da barra representa a quantidade de dias sem acesso ao moodle e a barra vertical representa o tempo a que se refere o dado, nesse caso partindo do ano.

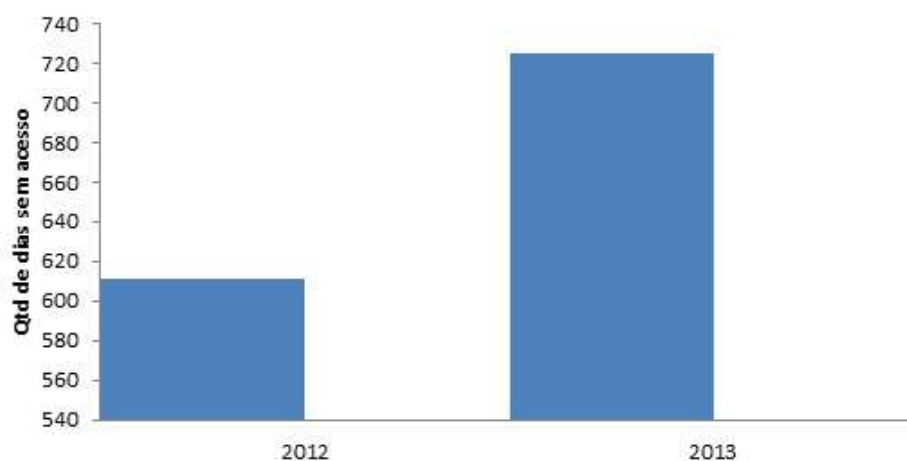
Figura 19 – Quantidade de dias sem acesso por ano

Ano	Dias_Sem_Acesso

Fonte: Elaborada pelo Autor.

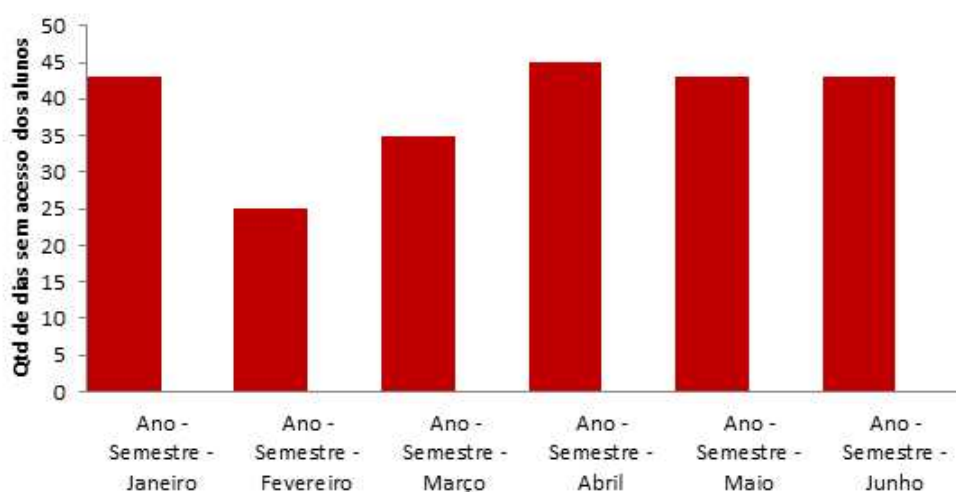
Como foi dito anteriormente, essas visões são pontos de partida para outras diversas possibilidade de pesquisa no cubo. Por exemplo, no gráfico acima parte-se de dados relacionado ao ano, podendo ocorrer um Drill Down para mostrar um dado mais específico como os dias sem acessos dos meses do primeiro semestre de um determinado ano, como representado no gráfico da Figura 21.

Figura 20 – Quantidade de dias sem acesso por ano.



Fonte: Elaborada pelo Autor.

Figura 21 – Quantidade de dias sem acesso por mês do primeiro semestre do ano.



Fonte: Elaborada pelo Autor.

3.4.5 Quantidade de dias que os alunos ficaram sem acessar o portal por curso

E finalizando, esta visão indica a quantidade de dias sem acessos dos alunos por curso em determinado ano, conforme mostra a tabela da Figura 22.

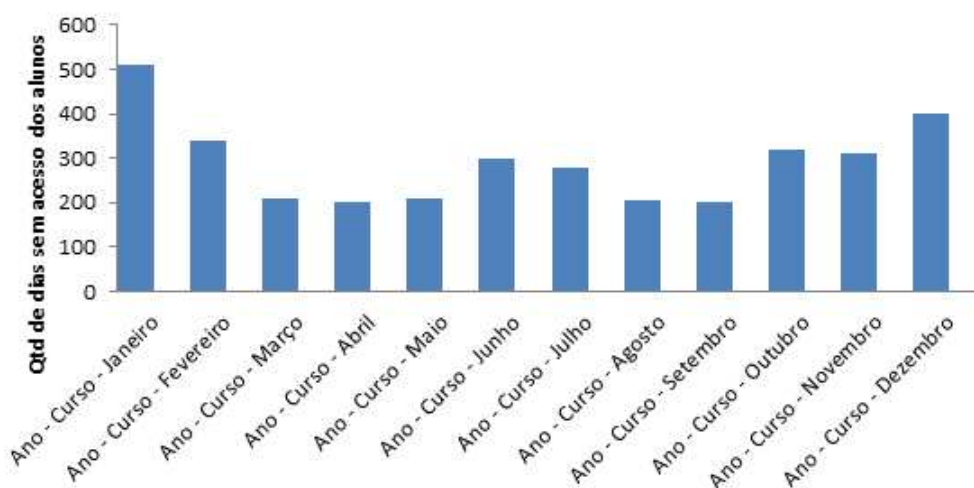
Figura 22 – Quantidade de dias sem acesso por curso e pelo tempo.

Ano	Mes	Curso	Dias_Sem_Acesso

Fonte: Elaborada pelo Autor.

Foi especificado um gráfico em barras onde a barra representa o tempo e o curso, e a sua altura representa a quantidade de dias sem acesso ao portal, representado no gráfico da Figura 23.

Figura 23 – Quantidade de dias sem acesso por curso e por mês.



Fonte: Elaborada pelo Autor.

Finalizando esta etapa de especificação das visões analíticas, encerra-se o modelo proposto por este trabalho, finalizando o capítulo com a validação e testes, seguindo com as considerações finais.

3.5 Validação e Testes

A própria criação da ferramenta OLAP com suas visões analíticas já é uma forma de validar o modelo, a partir dela já é possível verificar a realidade dos dados. Para isso foram feitas consultas na base de dados e os resultados foram comparados com os dados gerados nas visões analíticas a fim de validar o modelo proposto neste trabalho.

O fato de se criar efetivamente um ambiente analítico já valida esse modelo. O uso das visões é imprescindível para auxiliar os professores na avaliação processual, sem elas fica muito difícil de identificar se os alunos estão realmente participando e aprendendo o

conteúdo pedagógico, visto que, o moodle é uma ferramenta que não tem um propósito analítico. Por exemplo, os dados referentes ao *quiz*, disponibilizados para os professores são bem pontuais, como mostra na Figura 24, onde conta com o nome do aluno, o dia de início, de finalização, tempo gasto e a sua avaliação. Para uma análise desses dados para identificar quantos alunos de uma disciplina não concluíram o *quiz*, por exemplo, seria necessário exporta-lo para uma estrutura auxiliar para posterior análise, dados esses de uma disciplina específica, sendo necessário repetir o procedimento para outras disciplinas desejadas (GOTTARDO, 2012).

Figura 24 – Página de acompanhamento do quiz.

Nome / Sobrenome	Iniciado em	Completo	Tempo utilizado	Avaliar/5	#1	#2	#3	#4	#5	#6	#7	#8	#9
	15 outubro 2014, 18:53	15 outubro 2014, 19:33	39 minutos 34 segundos	2,78	0/0,56	0,56/0,56	0,56/0,56	0,56/0,56	0/0,56	0,56/0,56	0/0,56	0/0,56	0,56/0,56
	15 outubro 2014, 19:37	15 outubro 2014, 19:37	13 segundos	5	0,56/0,56	0,56/0,56	0,56/0,56	0,56/0,56	0,56/0,56	0,56/0,56	0,56/0,56	0,56/0,56	0,56/0,56
	14 outubro 2014, 21:12	14 outubro 2014, 21:43	30 minutos 12 segundos	4,44	0,56/0,56	0,39/0,56	0,56/0,56	0,56/0,56	0,56/0,56	0,44/0,56	0,5/0,56	0,33/0,56	0,56/0,56
	14 outubro 2014, 21:44	14 outubro 2014, 21:51	6 minutos 35 segundos	5	0,56/0,56	0,56/0,56	0,56/0,56	0,56/0,56	0,56/0,56	0,56/0,56	0,56/0,56	0,56/0,56	0,56/0,56
	14 outubro 2014, 21:52	14 outubro 2014, 21:56	3 minutos 53 segundos	5	0,56/0,56	0,56/0,56	0,56/0,56	0,56/0,56	0,56/0,56	0,56/0,56	0,56/0,56	0,56/0,56	0,56/0,56
	16 outubro 2014, 15:08	16 outubro 2014, 15:16	7 minutos 52 segundos	5	0,56/0,56	0,56/0,56	0,56/0,56	0,56/0,56	0,56/0,56	0,56/0,56	0,56/0,56	0,56/0,56	0,56/0,56
	11 novembro 2014, 21:58	11 novembro 2014, 22:33	35 minutos 26 segundos	5	0,56/0,56	0,56/0,56	0,56/0,56	0,56/0,56	0,56/0,56	0,56/0,56	0,56/0,56	0,56/0,56	0,56/0,56

Fonte: Moodle.

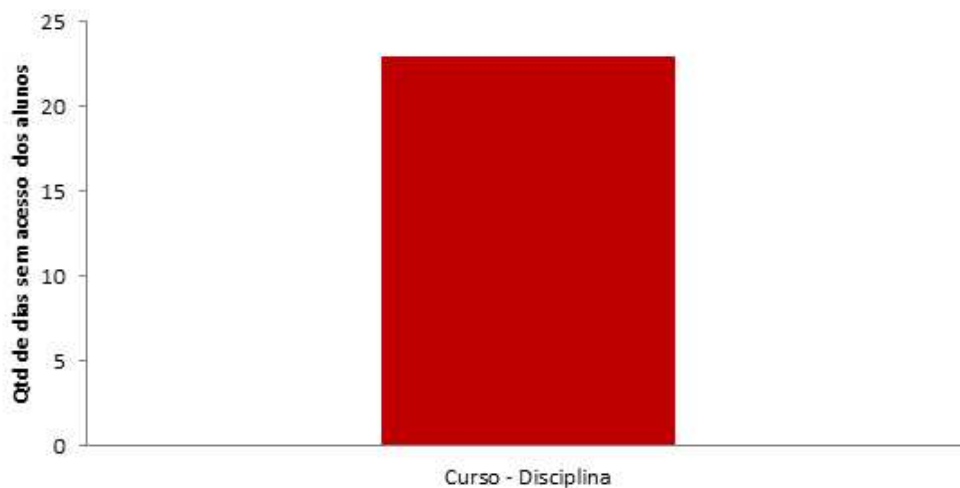
Nas visões analíticas criadas neste trabalho basta clicar na dimensão disciplina e arrastar para área de filtro que abrirá uma tela com as disciplinas a serem selecionadas, prontamente os dados estarão disponíveis em forma de tabela ou gráfico, conforme na Figura 25.

Assim como na atividade de *quiz*, a informação que o professor tem com relação a frequência dos alunos é muito específica e de difícil análise. O cenário encontrado pelo professor no moodle é o mostrado na Figura 26, onde tem as seguintes informações: a foto e o nome do aluno, sua cidade e país, e o último acesso do mesmo. Da mesma forma que na atividade analisada anteriormente, o dado para ser tratado teria que ser extraído para uma estrutura auxiliar para análise.

A quarta visão apresentada já dá como ponto de partida a quantidade de dias sem acesso por ano, podendo obter essa informação de dados mais específicos através de uma operação de Drill Down dando um duplo click no gráfico, que passará a mostrar os dados com relação ao semestre, conforme já representado na Figura 21.

Com o modelo finalizado e validado, o próximo capítulo faz as considerações finais.

Figura 25 – Quantidade de dias sem acesso por disciplina.



Fonte: Elaborada pelo Autor.

Figura 26 – Página de acompanhamento de acesso.

Foto do usuário	Nome / Sobrenome	Cidade/Município	País	Último acesso ↑	Selecionar
		Lauro de Freitas	Brasil	14 horas 22 minutos	<input type="checkbox"/>
		Lauro de Freitas	Brasil	14 horas 44 minutos	<input type="checkbox"/>
		Salvador	Brasil	20 horas 10 minutos	<input type="checkbox"/>
		Salvador	Brasil	20 horas 55 minutos	<input type="checkbox"/>
		Brumado	Brasil	1 dia	<input type="checkbox"/>
		Simoes Filho	Brasil	1 dia 20 horas	<input type="checkbox"/>
		simões filho	Brasil	3 dias 14 horas	<input type="checkbox"/>
		Lauro de Freitas	Brasil	6 dias 18 horas	<input type="checkbox"/>
		Simoes Filho	Brasil	6 dias 19 horas	<input type="checkbox"/>

Fonte: Moodle.

4 Considerações Finais

O provimento de informações pelo Business Intelligence para auxiliar gestores já é uma realidade. Os bancos de dados operacionais camuflam muitas informações importantes que podem servir de indicadores para um melhor gerenciamento. No entanto, ainda é pouco utilizado em áreas como da educação, principalmente na área da educação à distância, que tem a difícil missão de tornar possível a avaliação processual num curso que tem como principal característica a comunicação não presencial, feita por meio das mídias.

Com o objetivo de auxiliar os professores na avaliação processual no ensino à distância, surgiu o desafio desta pesquisa de criar um modelo de Business Intelligence para análise de desempenho dos alunos de um curso EAD. Este modelo é constituído de uma arquitetura técnica, uma arquitetura de BI e algumas visões analíticas.

Este projeto foi motivado pela possibilidade de contribuir com a difícil missão de acompanhar o desempenho dos alunos de uma instituição de educação à distância. Contribuir com a análise de dados importantes que estão dispersos na base de dados do moodle, portal usado como meio entre alunos, professores e suas atividades.

A construção deste projeto contou com uma análise profunda da fonte de dados para o entendimento do mesmo, a instalação do moodle, do banco de dados usado e um estudo da sua estrutura. Fase complexa, visto que o banco conta com 199 tabelas de difícil entendimento e pouca documentação. Após essa fase de entendimento da fonte de dados, foi iniciado a construção do *Data Warehouse*, onde foram criados dois *Data marts*, um relacionado a frequência com que os alunos acessam o moodle e o outro referente a atividade de *Quiz* identificando o andamento da mesma. Para a criação dos *Data Marts* foi utilizada a metodologia bottom-up, também conhecida como Star-Schema, proposta por Kimball, e que se mostra uma metodologia simples e com um retorno mais rápido. Um abordagem caracterizada por projetos pequenos e independentes, focado em assuntos específicos com o intuito de reduzir a complexidade.(SANSU, 2012).

Após a criação do DW, este projeto contou com a criação de um ETL, processo que extrai os dados da fonte, transformam esses dados a fim de adequá-los ao *Data Warehouse* e por fim carrega-los. Com o DW criado e os dados carregados, foi utilizado o Pentaho como ferramenta OLAP para mapear os cubos e gerar algumas visões analíticas para servir de ponto de partida para os professores, não substituindo a livre exploração do cubo de dados. A criação das visões analíticas foi de extrema importância, pois foram feitas comparações, com as mesmas condições, dos dados gerados nas visões com dados gerados em consultas feitas diretamente no banco de dados, validando as informações que

estavam sendo mostradas.

Esta pesquisa encontrou algumas dificuldades essencialmente no entendimento da fonte de dados, pois é uma base de dados extensa e não contava com uma documentação oficial que desse um apoio para maior esclarecimento.

Os objetivos traçados para este projeto foram cumpridos, este trabalho gerou indicadores que auxiliam os professores na avaliação processual de seus alunos. Por se tratar de uma monografia não foram abrangidas todas as atividades que são comumente utilizadas num curso EAD, o objetivo deste trabalho foi de mostrar que é possível o uso do BI para esse ramo da educação.

Por fim, sugere-se para a continuação deste trabalho o uso de outras ferramentas OLAP do mercado e um estudo sobre quais informações de um curso EAD são mais importantes para uma melhor avaliação processual.

Referências

- ALVES, L.; NOVA, C. *Educacao a distancia: limites e possibilidades*. Sao Paulo: [s.n.], 2003. Citado na página 15.
- BARBIERI, C. *BI2 - Business Intelligence: Modelagem & qualidade*. Rio de Janeiro: Campus, 2011. Citado 3 vezes nas páginas 11, 20 e 24.
- GEORGE, S. *Inmon vs. Kimball: Which approach is suitable for your data warehouse?* 2012. Acesso em: 27 de novembro de 2014. Disponível em: <<http://searchbusinessintelligence.techtarget.in/tip/Inmon-vs-Kimball-Which-approach-is-suitable-for-your-data-warehouse>>. Citado na página 20.
- GOTTARDO, E. *Previsao de Desempenho de Estudantes em Cursos EAD Utilizando Mineracao de Dados: uma Estrategia Baseada em Series Temporais*. Rio Grande do Sul: [s.n.], 2012. Citado na página 41.
- INMON, W. *Building the Data Warehouse*. New York: [s.n.], 2002. Citado 2 vezes nas páginas 18 e 20.
- KIMBALL, R. *The Data Warehouse Toolkit: Guia Completo Para Modelagem Dimensional*. New York: Campus, 2002. Citado 6 vezes nas páginas 18, 20, 21, 27, 28 e 31.
- MONDRIAN, C. P. *Mondrian*. 2005. Acesso em: 5 de dezembro de 2014. Disponível em: <<http://community.pentaho.com/projects/mondrian/>>. Citado na página 30.
- PASCHOAL, J. P. *Modelo de Business Intelligence Para Pequenos Mercados*. Salvador: [s.n.], 2012. Citado na página 24.
- PENTAHO, C. *Pentaho Solutions*. 2005. Acesso em: 27 de novembro de 2014. Disponível em: <<http://www.pentaho.com/>>. Citado na página 30.
- PMI, P. M. I. *PMI Um guia do conhecimento em gerenciamento de projetos (PMBok)*. [S.l.: s.n.], 2008. Citado na página 26.
- SETZER, V. *Os Meios Eletronicos e a Educacao: Uma Visao alternativa*. Sao Paulo: [s.n.], 2001. Citado 2 vezes nas páginas 13 e 14.
- SEZOES, C. *Manual Business Intelligence*. PORTO: [s.n.], 2006. Citado na página 16.
- VERCELLIS, C. *Data Mining and Optimization for Decision Making*. Chichester: [s.n.], 2009. Citado 2 vezes nas páginas 16 e 24.