

Data Warehouse: Aplicação e usabilidade em um cenário realístico

Lucas J. Martins¹
Nathan S. Werlich¹
Thibault Fruchart²
Wesley M. Luz¹

Instituto de Informática e Estatística – INE

¹Universidade Federal de Santa Catarina (UFSC)
Florianópolis – SC – Brasil

²Institut Supérieur d'Electronique et du Numérique (ISEN)
Lille – França

nathanwerlich@gmail.com, lucasjoao.lj@gmail.com, wmayk007@gmail.com,
thibault.fruchart@isen.yncrea.fr

Abstract. *This article describes the experiences and learning gained through the practical implementation of a Data Warehouse / Data Mart, as well as a specific real scenario and its demand for information. Such a scenario is based on the need for a pre-college course in expanding its network, as well as reaching the right customers. The source of information for such appointments and surveys are Coperve-UFSC vestibular data from 2008 to 2012.*

Resumo. *Este artigo descreve as experiências e aprendizados obtidos através da implementação prática de um Data Warehouse/Data Mart, e também um cenário real específico e sua demanda por informações. Tal cenário baseia-se sobre a necessidade de um cursinho pré-vestibular em expandir sua rede, assim como alcançar os clientes corretos. A fonte de informações para tais apontamentos e pesquisas são os dados de vestibulares da Coperve-UFSC ocorridos entre 2008 a 2012.*

1. Introdução

A partir de proposta de trabalho final da disciplina INE5643 - Data Warehouse na universidade Federal de Santa Catarina, surgiu a necessidade de criação de um cenário realístico para implantação e desenvolvimento de toda a estrutura associada a um DataWarehouse/DataMart como forma de verificação de aprendizado.

Tal projeto consiste em colocar em prática todo o conhecimento adquirido ao longo do semestre, de modo a integrar as diferentes etapas do projeto de um Data Warehouse e evidenciar os possíveis resultados e ganhos que podem ser obtidos através da utilização do mesmo. Através de definições de requisitos e necessidades de um possível cliente, utilização de ferramentas de modelagem para diagramas, ferramentas para desenvolvimento de back end, front end e elaboração de um artigo, espera-se que haja uma síntese de todos os conteúdos abordados em sala de aula sobre Data Warehouse, processos ETL/OLTP e BI.

Como abstração de um cenário real que demanda a necessidade de implantação de um Data Mart/Data Warehouse para obtenção de resultados e respostas específicas, foi escolhido atender a um cliente fictício, possuidor de uma rede de cursinhos pré-vestibular, que pretende expandir sua rede de ensino, assim como encontrar os clientes específicos que mais se encaixam em seu público alvo. De modo a direcionar recursos e propagandas à áreas com maior incidência de seu seu público alvo, resultando no aumento de alunos do cursinho, aumento do lucro e resultados gerais de aprovação e expansão da rede.

2. Materiais

Através de dados disponibilizados acerca dos vestibulares da UFSC, entre os anos de 2008 à 2012, realizados pela Coperve buscou-se encontrar informações e indicadores úteis, que ilustrassem novos conhecimentos acerca dos vestibulandos. Entretanto, conhecimentos estes que venham a ser do interesse do domínio de nosso problema. É do interesse do escopo de uma rede de cursinhos pré-vestibulares a obtenção de respostas como:

- Porcentagem de candidatos aprovados em cursos específicos e/ou áreas de conhecimento específicas, separados por mesorregiões do estado;

- Quantidade de reprovados de regiões específicas que foram demasiadamente mal em uma determinada matéria/área de conhecimento, agrupados por faixas salariais;

- Informações específicas de aprovados ou reprovados em cursos com alta demanda por vaga, separados por região em vestibulares/anos específicos;

Foi entendido que, através da implantação e desenvolvimento de um Data Mart que responda a tal demanda de informações, será possível atender as necessidades da Rede de Cursinhos pré-vestibulares, justificando assim o investimento em uma equipe de desenvolvimento de Data Warehouse.

3. Métodos

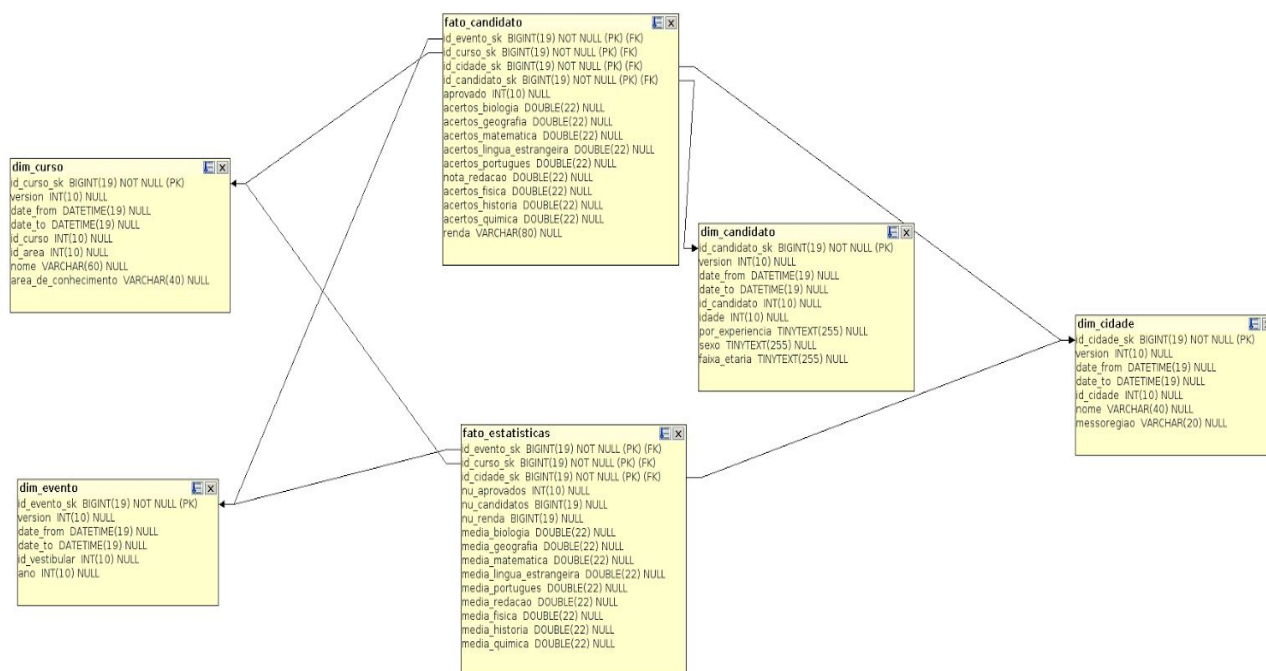
De modo geral, um Data Warehouse/Data Mart pode ser definido como uma tecnologia que armazena dados e informações detalhadas de um domínio de negócio, visando responder perguntas específicas, auxiliar na criação, organização e elaboração de relatórios que visam ajudar na tomada de decisões significativas da organização, tendo como base os fatos e indicadores apresentados pelo data warehouse.

Partindo desta ciência, fora decidido que esta tecnologia encaixa-se perfeitamente ao escopo de problema que necessita ser resolvido.

Para implantação/desenvolvimento do mesmo, foram utilizados como ferramentas de back end e front end, respectivamente, Kettle e Metabase.

4. Metodologia:

A partir dos requisitos do cliente, e das informações contidas no Banco de Dados a nós disponibilizado, foi possível abstrair e retirar aquilo que nos era necessário, para então formar um esquema estrela contendo as seguintes dimensões e fatos:



Em relação aos aspectos desenvolvidos no back end, através da ferramenta “Kettle” da Pentaho, destacam-se para exemplificação os pontos/transformações realizadas nas imagens a seguir:



A partir da data de nascimento de um determinado candidato, são realizadas operações com a data atual do sistema, para fornecer então, a idade correta do candidato.

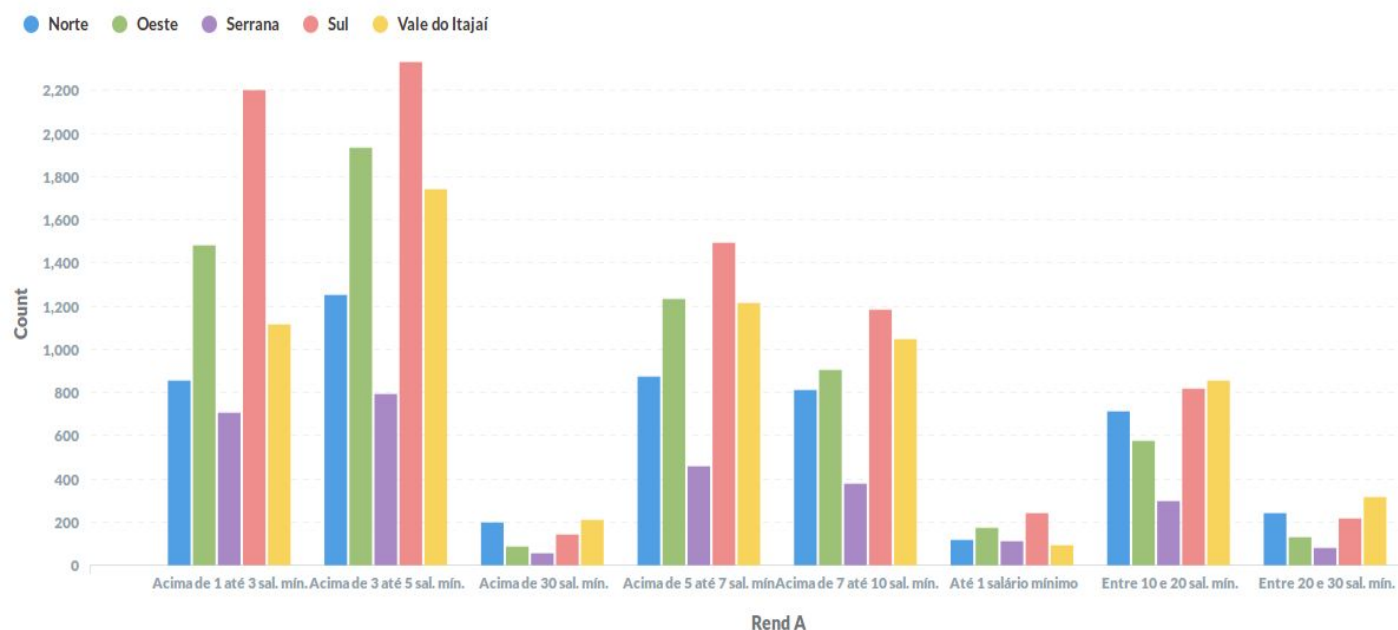


Através de uma lista de cidades catarinenses, uma planilha foi montada na ferramenta “Excel” com as suas respectivas mesorregiões de pertencimento. Então, tais informações são inseridas para associação no Kettle, juntamente com a criação de “id’s” para estas cidades.

5. Resultados:

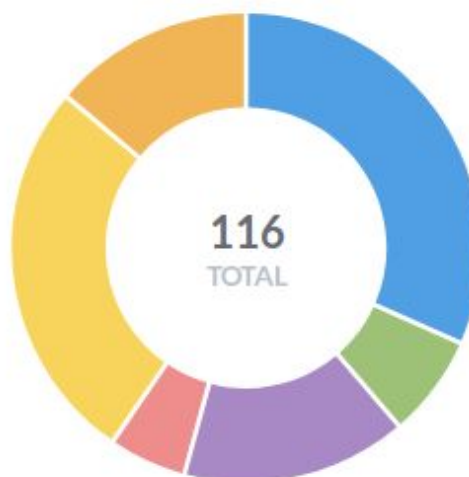
Os resultados obtidos, podem ser verificados e sintetizados através dos gráficos e amostragens obtidas na ferramenta “MetaBase”, que além de apresentar informações antes desconhecidas, também responde à questões cruciais do escopo de negócio do cliente.

Reprovados que tiraram < 6 em Física/Matemática que não residem na Grande Florianópolis



Aprovados em SI/CCO por mesoregião

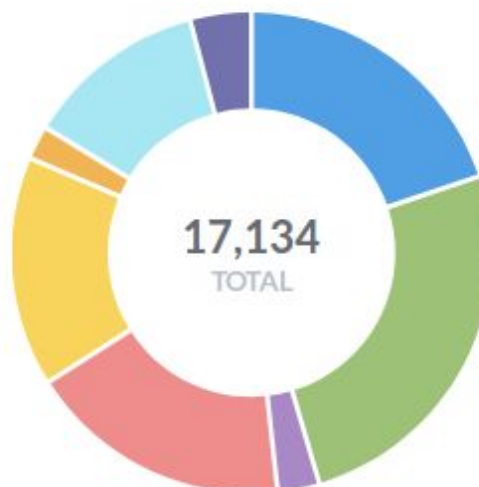
Grande Florianópolis	31.90%
Norte	6.90%
Oeste	15.52%
Serrana	5.17%
Sul	26.72%
Vale do Itajaí	13.79%



Conforme exemplificado nas imagens acima, foram obtidos indicadores/respostas acerca de aprovados ou reprovados em cursos com alta demanda por vaga, separados por região e/ou anos específicos. Ou, determinadas áreas de conhecimento/matérias em que vestibulandos obtiveram desempenho crítico.

Renda dos Reprovados da região Sul e Vale do Itajaí

Acima de 1 até 3 sal. mín.	20.02%
Acima de 3 até 5 sal. mín.	25.88%
Acima de 30 sal. mín.	2.60%
Acima de 5 até 7 sal. mín.	17.84%
Acima de 7 até 10 sal. mín.	15.50%
Até 1 salário mínimo	2.04%
Entre 10 e 20 sal. mín.	12.19%
Entre 20 e 30 sal. mín.	3.92%



Na imagem acima é possível verificar indicadores de renda dos candidatos que foram reprovados no vestibular, separado por mesorregião Sul e Vale do Itajaí.

Candidatos que tiraram < 5 em Historia/Geografia Separado por area de conhecimento

Mes So Regia O	Ciências Biológicas	Ciências Exatas	Ciências Sociais	
Grande Florianópolis	281	264	694	15
Norte	211	208	276	6
Não informado	3,660	3,333	3,872	79
Oeste	651	600	571	22
Serrana	190	156	167	6
Sul	427	445	574	19
Vale do Itajaí	407	355	547	23

Rows 1-7 of 8 ◀ ▶

Enquanto que, na imagem acima, o critério de amostragem é baseado em candidatos com nota menor que 5 em determinadas matérias, agrupados por mesorregiões e áreas de conhecimento.

6. Conclusão e Trabalhos Futuros:

Futuramente, pretende-se expandir as análises para outras regiões, de modo que, a rede de cursinhos pré-vestibulares possa estender seus serviços oferecidos à outros estados e buscar novos indicadores/questões a serem respondidas, cada vez mais significativos/as e exatos, de modo a auxiliar diretamente na tomada de decisões. Além disto, pretende-se também utilizar a base de dados com dados de outros vestibulares em diversas universidades públicas.

De modo geral, é possível afirmar que durante o desenvolvimento e implantação do projeto final de tal disciplina, foi possível a sintetização de todos os conhecimentos obtidos ao longo do semestre, paralelamente à assimilação de novos conhecimentos e experiências. Tornando-se algo de extrema importância para o desenvolvimento técnico e científico do grupo como um todo, e proporcionando aprendizado efetivo e duradouro.

7. Bibliografia:

<https://canaltech.com.br/business-intelligence/conhecendo-a-arquitetura-de-data-warehouse-19266/>

Autor: Diego Elias, em 08 de Abril de 2014. Acessado em: Novembro, 2017.

[https://wiki.pentaho.com/display/EAI/Pentaho+Data+Integration+\(Kettle\)+Tutorial](https://wiki.pentaho.com/display/EAI/Pentaho+Data+Integration+(Kettle)+Tutorial)

Acessado em: Novembro, 2017.

<https://www.metabase.com/blog/how-to-use-metabase/>

Acessado em: Novembro, 2017.