

Apoio à PRAE (moradia estudantil) e setor imobiliário da grande Florianópolis quanto a demanda por moradia dos candidatos aprovados de fora da grande Florianópolis

Caio Franco Suarez dos Santos, Lucas Pagotto Tonussi, Rafael Barbaresco

Curso de Bacharelado em Sistemas de Informação

Universidade Federal de Santa Catarina (UFSC)

Prof. Dr. José Leomar Todesco

caiosasia@gmail.com, lptonussi@gmail.com, rafaelbrbse@gmail.com, jose.todesco@ufsc.br

Resumo. Este projeto tem o objetivo de analisar o modelo sócio acadêmico do vestibular da Coperve de 2008 a 2012, com a finalidade de implementar um Data Mart para dar suporte a análises sobre o tema “Apoio à PRAE (moradia estudantil) e setor imobiliário da grande Florianópolis quanto a demanda por moradia dos candidatos aprovados de fora da grande Florianópolis”. Para isso foram formuladas algumas perguntas norteadoras que serão respondidas por meio da modelagem, tais como: Anualmente, qual a quantidade de estudantes de fora da grande Florianópolis de renda baixa separados por cidades que foram aprovados? Anualmente, qual a quantidade de estudantes de fora da grande Florianópolis de renda média ou alta separados por cidades que foram aprovados?. Para realizar a modelagem foi disponibilizado pela Coperve um banco de dados com dados dos vestibulares de 2008 a 2012. Além disso foi feito uso do software Pentaho para a realização do processo de ETL e o Metabase para implementar a parte do front-end

Abstract. Paper describing the building of a data-warehouse for the aid of the PRAE system (a.k.a. Student Boarding Houses, Student Dormitories).

1. Introdução

Segundo dados da Coperve, em 2020 foram ofertadas mais de 4500 vagas no vestibular da ufsc, dentre essas vagas, 68% eram no campus de Florianópolis, que conta com 78 opções diferentes de curso. Por ser uma instituição renome, anualmente a ufsc recebe candidatos de diversas cidades, que vem fazer o vestibular buscando realizar o sonho de estudar nessa universidade.

Sendo assim, muitos dos estudantes aprovados anualmente não são moradores de Florianópolis e precisam de um lugar para morar durante os anos da graduação. Quem se beneficia com isso é o setor imobiliário que acaba encontrando uma grande clientela pelos arredores da UFSC. Porém nem todos os estudantes aprovados no vestibular tem condições de se mudarem para Florianópolis e pagar aluguel. Por isso todos os anos a Pró-Reitoria de Assuntos Estudantis (PRAE) disponibiliza vagas na moradia estudantil, para estudantes da universidade oriundos de famílias com renda bruta familiar de até 1,5 salário mínimo per capita.

Dessa forma, este projeto permitirá que tanto a PRAE quanto empresas do setor imobiliário, possam realizar um estudo aprofundado sobre a demanda por moradia dos candidatos aprovados de fora da grande Florianópolis, possibilitando uma preparação melhor para os próximos anos.

2. Embasamento Teórico

A seguir será exposto um breve embasamento teórico, para poder dar direcionamento, e auxílio, (ao leitor), mínimo sobre o que este trabalho está tratando.

2.1 Data Warehouse

Segundo Date (2004), Data Warehouse é um banco de dados orientado por assunto, integrado, não volátil, variável com o tempo, para apoiar as decisões da alta administração de uma organização. Kimball complementa essa definição, expondo que um DWH é “uma cópia de dados transacionais especialmente estruturada para query e análise.” Essas definições endereçam uma solução para atender aos funcionários da PRAE UFSC. A missão é modelar, a partir de um banco existente e construído via modelagem MER, uma derivação do banco de dados atual, para futuras análises mais apuradas. A seguir será explicado sobre como essa derivação pode ser feita, seu nome e do que se trata.

2.1 Modelo Estrela

A “derivação,” mencionada anteriormente, chama-se esquema estrela. Essa forma de modelagem se dá por uma estrutura simples, composta de várias tabelas dimensão e uma tabela fato. (POE, KLAUER, BROBST, 1998). As dimensões contém colunas descritivas / categóricas, cada dimensão tem uma chave artificial (será explicada adiante). As dimensões são de pouco volume e combinadas de formas diferentes elas explicam os fatos na tabela fato. A tabela fato deve conter as chaves estrangeiras, uma para cada dimensão, e colunas métricas, valoradas, devendo-se evitar colunas booleanas. As dimensões criadas para este trabalho são enxutas e divididas por assunto, para que não haja confusão entres mesmas. O foco principal na construção das

dimensões foi fornecer para os funcionários PRAE UFSC o mínimo de descrições possível para ajudá-los a decidir sobre a questão das moradias estudantis.

2.2 Chave artificial

A chave artificial é bastante simples, também chamada de chave surrogate. O criador do DWH pode definir a surrogate a bel-prazer contanto que não seja a mesma chave do modelo relacional fonte que fora usado para derivar a dimensão. Cada uma das dimensões criadas para o DWH deste trabalho se trata de uma chave primária auto incremental do tipo inteira, para facilitar a criação de chaves estrangeiras na tabela fato.

3. Materiais

A PRAE é um órgão executivo central integrante da Administração Superior da Universidade Federal de Santa Catarina. O objetivo da PRAE é auxiliar o Reitor em suas tarefas executivas, especialmente em programas e projetos voltados à política estudantil.

Dentre as atribuições da PRAE, compete a ela:

- Propor e acompanhar a execução de ações da política de assuntos estudantis da universidade, principalmente no que se refere ao acesso, à permanência e a conclusão do curso de graduação presencial. Dentre as áreas aplicáveis, inclui moradia estudantil;
- Estimular a implementação de planos, programas e projetos junto à comunidade estudantil;
- Propor e desenvolver políticas de benefícios da Universidade, dirigida à comunidade estudantil;

A PRAE oferece diversos benefícios, com vagas limitadas, aos estudantes que necessitam apoio financeiro. Para selecionar os estudantes que ocuparão as vagas, a PRAE possui um cadastro socioeconômico, que categoriza os estudantes mais necessitados de acordo com sua renda, tamanho da família, meio de transporte, entre outros dados. Para projetos de longo prazo de implementação, como construção de moradia estudantil, é de interesse da PRAE receber dados estatísticos de candidatos ingressantes nos vestibulares.

A Comissão Permanente do Vestibular (Coperve) é o setor administrativo da UFSC que gerencia o processo de elaboração, aplicação e correção das provas do vestibular. As provas são compostas por 80 questões de múltipla escolha distribuídas entre as disciplinas exigidas, uma redação, e até 6 questões discursivas, dependendo do ano de aplicação. A Coperve possui registro de todos os candidatos a vestibulares, o que inclui local de origem de cada candidato e desempenho individual.

A Coperve disponibilizou uma amostra do banco de dados com candidatos anônimos dos vestibulares de 2008 até 2012, bem como o cadastro socioeconômico preenchido por cada um deles. Essa amostra serve como base para elaboração de um projeto que oferece os dados estatísticos necessários para a PRAE, no contexto de construção de moradia estudantil, e serviços externos à UFSC do setor imobiliário.

4. Métodos

Utilizamos o método básico da literatura para construção do esquema estrela. O método consiste em [KIMBALL, 2013, pg. 1]:

1. Definição de quais processos de negócio está sendo modelado/tratado [1ª Etapa]:
 - a. Idealizamos modelar algo alinhado com “Anualmente qual é a quantidade de alunos que não moram na Grande Florianópolis E que prestaram vestibular, E passaram, E tem baixa renda?”
 - b. Vamos modelar quais estudantes passaram em qualquer curso que seja. O curso em si não importa, desde que o estudante passou.
 - c. Vamos modelar quais cidades os estudantes moram para mapear se eles precisam de moradia, pois se eles vem de outra cidade e tem uma renda baixa então precisam de moradia estudantil.
 - d. Precisamos saber informações sobre:
 - i. Os(as) candidatos(as) que passaram.
 - ii. Onde eles(as) moram.
 - iii. Para o PRAE poder escolher melhor sobre a questão da moradia
 1. Qual a nota dos candidatos(as)
 2. Qual a renda da família do candidato(a)
 3. Qual candidato tem meio de transporte próprio
2. Quais definições para os grãos do processo de negócio [2ª Etapa]:
 - a. Grão por:
 - i. candidatos(as) (pessoas) classificados(as)
 - ii. eventos dos vestibulares com temporalidade: anual
 - iii. cidades de onde os candidatos(as) vêm (morava inicialmente)
3. Quais dimensões serão aplicadas a cada registro da futura tabela de fatos [3ª Etapa]:
 - a. 5 dimensões foram descobertas, a partir do ER original da coperve / vestibular, adiante será explicado cada uma das dimensões.
4. Os fato(s) mensuráveis [4ª Etapa]:
 - a. 1 fato foi descoberto, adiante o mesmo será explicado.

5. Metodologia

Estabelecemos uma hipótese inicial para trabalhar em cima da modelagem do problema. A hipótese é de que o modelo ER da coperve para o vestibular poderia explicar a necessidade de expansão da necessidade de moradias estudantis ou não. Essa mesma dúvida fez surgir uma questão inicial “Anualmente, qual a quantidade de estudantes de fora da grande Florianópolis de renda média ou alta separados por cidades que foram aprovados?”, que foi carro chefe para estudarmos propostas. Para modelar o esquema estrela foi utilizado o princípio bottom-up proposto por Kimball, e também as 4 etapas para construção do esquema estrela [KIMBALL, 2008]. Também utilizamos de material disponível no sítio online do Grupo Kimball [KIMBALL, 2013]. Foram feitas reuniões diárias para discussão e modelagem do problema, questões de apoio para resolução do problema foram propostas e a partir delas foi possível começar a descobrir mais e mais sobre a hipótese inicial.

6. Resultados

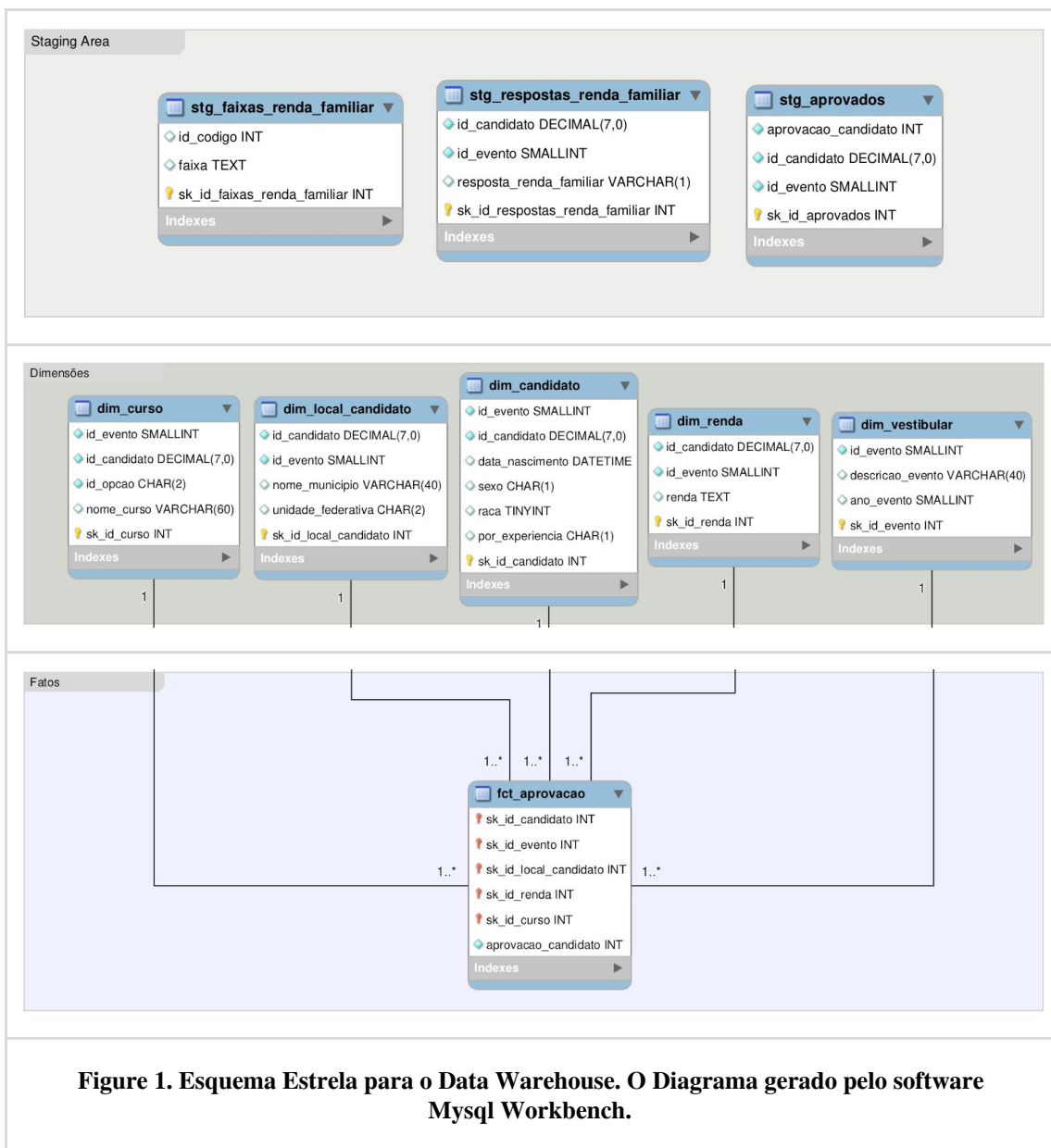
Após execução das 4 etapas de modelagem do esquema estrela, como fora descrito anteriormente, foi possível descobrir e gerar 5 dimensões para compor o fato final, as dimensões e o fato serão explicados mais adiante.

6.1. Dimensões descobertas

Dimensões Criadas	
Dimensão Localização Candidato Permite explicar a respeito de onde o candidato mora.	Dimensão renda Permite explicar sobre a escolha do candidato em relação a renda familiar
Dimensão candidato(a) Permite explicar um pouco sobre o(a) candidato(a)	Dimensão vestibular Permite explicar os fatos segmentando por vestibular (evento do concurso, grão tempo anual)
Dimensão curso Permite explicar sobre os cursos escolhidos pelos candidatos e segmentar quais cursos tem mais candidatos recebendo moradia. Pode-se filtrar por opção={1, 1A, 2}	
Tabelas Staging (Intermediárias)	
Foram criadas tabelas intermediárias para podermos estabelecer os dados no fato. stg_aprovados, stg_faixas_renda_familiar, stg_respostas_renda_familiar. O SQL dessas tabelas estará disponível em um repositório no github, disponível em < https://github.com/tonussi/dw-coperve/tree/master/dw/stg >.	
Tabela 1. Dimensões encontradas para endereçar o problema proposto. As mesmas foram definidas a partir do esquema ER que define os dados relevantes para os vestibulares UFSC (Fonte Coperve, Disponibilizado pelo Professor Prof. Dr. José Leomar Todesco e Prof. Dr. Olinto).	

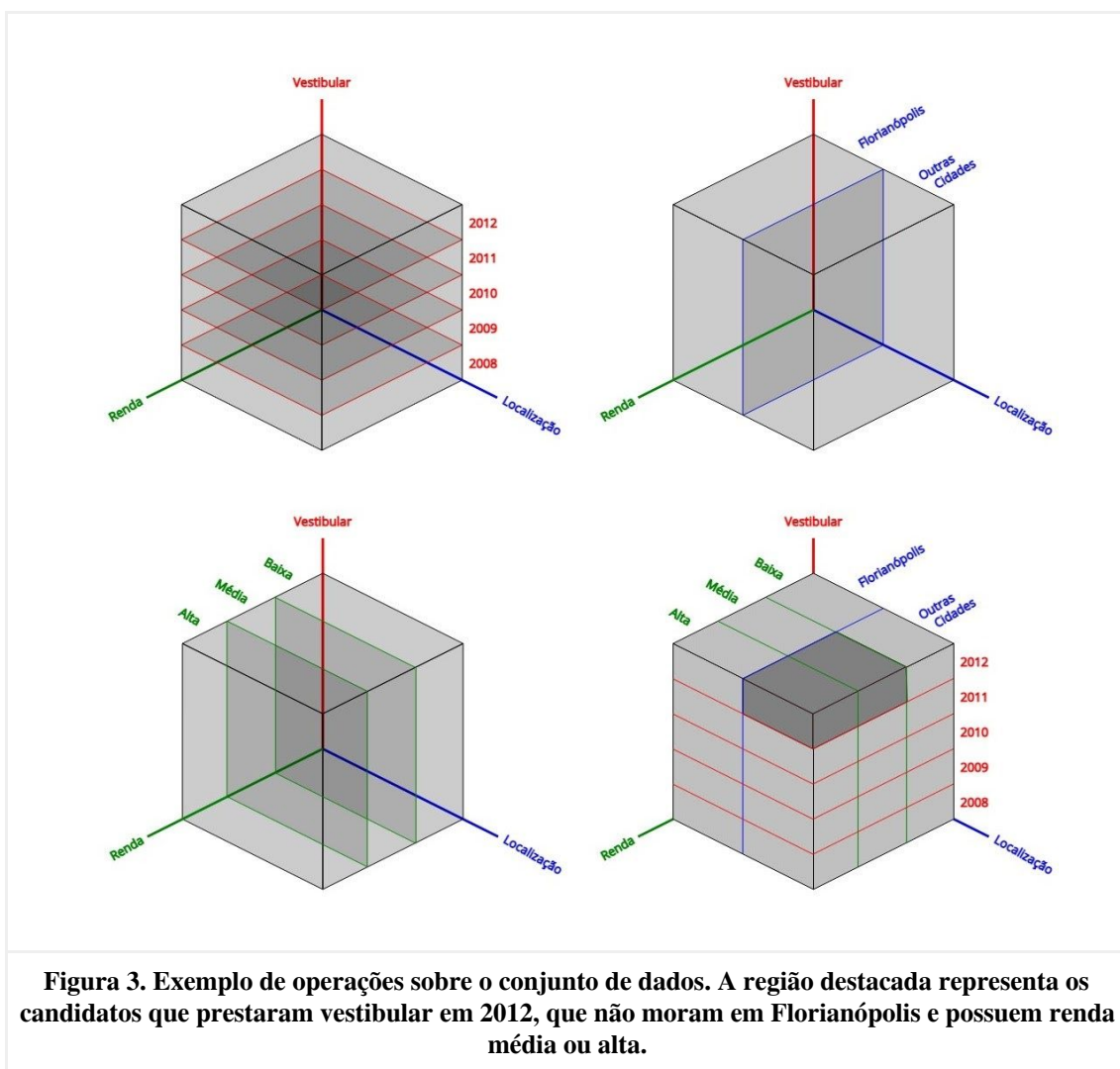
6.3. Fato descoberto

Foi gerado, a partir das dimensões, 1 fato apenas.



6.4 Representação Gráfica

O modelo estrela pode ser representado graficamente, desde que envolvam 3 ou menos dimensões. Considerando apenas as dimensões “Localização”, “Renda” e “Vestibular”, podemos representar o conjunto de elementos da tabela fato (“Aprovações”) contidos em um cubo. A figura 3 apresenta um exemplo de operações sobre o conjunto de dados para selecionar apenas os candidatos que prestaram vestibular em 2012, que não moram em Florianópolis e possuem renda média ou alta.



6.5. Possíveis riscos

Na construção da tabela fato observamos a repetição de muitos candidatos por conta da dimensão curso que contém 3 tipos de escolha para opção de curso a ser cursado se o candidato passasse no vestibular. O risco existe no frontend (Metabase) e para tratar esse problema foi necessário aplicar queries com “distinct” sobre os “*identificadores dos candidatos*”. Mais sobre o Frontend e Metabase será discutido a seguir.

Existem riscos relacionados às perguntas, gráficos, indicadores, métricas não atenderem à necessidade do PRAE de tratar o problema de moradias para os estudantes de baixa renda, de fora de Florianópolis. Para solucionar esse problema, seria necessário entrevistas com os funcionários do PRAE para entender as dores deles. Seria necessário capturar a necessidade através de reuniões com os superintendentes do PRAE.

Um dos riscos é o de interpretar mal o que é renda baixa, para esse trabalho consideramos incluir as respostas para renda como sendo 1-5 salários mínimos, sabe-se

que existe bastante dificuldade para a família de fora da Grande Florianópolis, em bancar o filho dentro de Florianópolis, onde tudo é muito caro.

O que pode-se fazer, dada as circunstâncias, é implantar o software, disponibilizar logins de acesso para o pessoal do PRAE e fazer uma avaliação, implantar um sistema de *Customer Success*, e fazer o possível para atender as necessidades dos funcionários do PRAE, e aceitar sugestões de mudanças.

Como usamos o FREE TIER da AWS para instanciar a entrega do BI para o PRAE UFSC, temos certeza que as máquinas e o banco de dados não vão aguentar muitas requisições diárias, causando gargalo na entrega dos resultados, pela ferramenta WEB de BI. Para solucionar esse risco, basta aumentar a potência das máquinas e a potência do AWS RDS Mysql. Aumentar o número de máquina, adicionando Auto Balanceamento de Carga. Assim já será suficiente atender a demanda do PRAE.

7. Implantação

Na implantação optamos pelo uso dos serviços provisionados pela AMAZON AWS Elastic Beanstalk (EBS). Esse serviço se mostrou prático e mais vantajoso na questão de “entrega contínua rápida” do que outros meios que a própria AWS oferece (i.e. Levantar uma máquina IAM personalizada e configurá-la do zero).

O EBS oferece uma pipeline de configurações automáticas que permitem entregar uma aplicação web, em poucas horas. No caso a aplicação web se trata do Metabase. O Metabase é uma forma de construir um BI integrado ao DWH. Neste mesmo, é possível com que a equipe de desenvolvedores em conjunto com outras áreas (i.e. outras engenharias, ciências econômicas, etc) possam construir a apresentação dos dados de forma ágil.

Passos da Implantação:

1. Criação Elastic Beanstalk com Banco de Dados Mysql Integrado
2. Seguir o guia oficial do Metabase para implantação na AWS, disponível em: [<https://www.metabase.com/docs/latest/operations-guide/running-metabase-on-elastic-beanstalk.html>](https://www.metabase.com/docs/latest/operations-guide/running-metabase-on-elastic-beanstalk.html).
3. Acessar o Banco de Dados Mysql criado pela AWS e:
 - 3.1. Construir a base de dados do vestibular
 - 3.2. Construir as tabelas de Staging
 - 3.3. Construir as dimensões
 - 3.4. Construir os fatos
 - 3.5. Popular os fatos

Conta de acesso, do professor, ao Metabase (BI), e também a URL de acesso.

ID: jose.todesco@ufsc.br

Senha: ###

[<http://metabaseprae-env.eba-aszkk3vh.us-east-1.elasticbeanstalk.com/>](http://metabaseprae-env.eba-aszkk3vh.us-east-1.elasticbeanstalk.com/)

8. Conclusões e Trabalhos Futuros

Em teoria para poder ser concluído esse trabalho deveria haver uma etapa de avaliação da ferramenta sendo utilizada pelos superintendentes do PRAE UFSC, só assim poderíamos coletar dados do uso, feedbacks, críticas, e eventuais problemas pós implantação.

8.1 Perguntas Criadas

Algumas questões para auxiliar os superintendentes da PRAE foram criadas, e no Metabase > Dashboard é possível encontrar as respostas para essas perguntas, com auxiliares visuais (i.e. gráficos).

1. Anualmente, qual a quantidade de estudantes de fora da grande Florianópolis de renda baixa separados por cidades que foram aprovados? (útil para prae)
2. Anualmente, qual a quantidade de estudantes de fora da grande Florianópolis de renda média ou alta separados por cidades que foram aprovados? (útil para setor imobiliário)
3. Anualmente, qual a quantidade de estudantes de fora da grande Florianópolis de renda baixa separados por cidades que foram reprovados? (que precisam de moradia mas não foram aprovados)
4. De quais cursos são os candidatos de baixa renda aprovados? Com escolha de opções={1, 1A, 2} (cursos escolhidos)

8.2 Aprendizados

Aprendemos muito sobre a transformação de dados de um MER fonte para um DWH (star-schema) que possibilita a construção de perguntas-respostas, indicadores, etc.

Aprendemos bastante sobre como a implantação de um sistema real, na AWS Amazon é complicado, e por vezes demorado, visto que o sistema deverá atender muito clientes, acessos simultâneos, e é necessário configurar uma infra-estrutura mínima e que tenha desempenho adequado, sem deixar o usuário esperando muitos minutos para obter suas respostas.

Bibliografia

- KIMBALL, R., Ross, M., THORTHWAITE, W., BECKER, B., MUNDY, J. (2008). **The data warehouse lifecycle toolkit**. Indianapolis, IN: Wiley, 2a edição. Acessado em: 12 de Novembro de 2020.
- KIMBALL, G. (2013). **Dimensional Modeling Techniques**. Disponível em: <<http://www.kimballgroup.com/wp-content/uploads/2013/08/2013.09-Kimball-Dimensional-Modeling-Techniques11.pdf>>. Acesso em: <21 de Novembro de 2020>. Acessado em: 13 de Novembro de 2020.
- DATE, C. J.. **Introdução A Sistemas de Banco de Dados**. 8ª Ed. Rio de Janeiro Campus, 2004. Acessado em: 24 de Novembro de 2020.
- POE, V., KLAUER, P., BROBST, S.. **Building a Data Warehouse for Decision Support**. ISBN: 9780137696390. Prentice Hall PTR. Disponível em: <<https://books.google.com.br/books?id=KJRRAAAAMAAJ>>. 1998. Acessado em: 25 de Novembro de 2020.