

Associação Propagadora Esdeva
Centro Universitário Academia –
UniAcademia Curso de Engenharia de
Software
Trabalho de Conclusão de Curso – Artigo

O USO DO BUSINESS INTELLIGENCE PARA ANÁLISE DE DADOS SOBRE COVID-19

Arthur Lourenço Terra¹

Centro Universitário Academia, Juiz de Fora, MG

Evaldo de Oliveira da Silva²

Centro Universitário Academia, Juiz de Fora, MG

Linha de Pesquisa: Business Intelligence

RESUMO

Este artigo consiste na construção de uma aplicação de *Business Intelligence* para permitir a análise de indicadores na área da saúde, tendo como fonte os dados a respeito do vírus COVID-19. Os dados foram extraídos de sites do Ministério da Saúde e da OMS (Organização Mundial da Saúde). Os indicadores foram especificados por profissionais da área da saúde a partir da necessidade de obterem informações úteis sobre a evolução do número de casos do COVID-19 na região de Juiz de Fora e outras regiões. Para armazenamento dos dados gerados para os indicadores, este trabalho apresenta as etapas do ETL (*Extract, Transform and Load*) implementado por meio da ferramenta *Pentaho Data Integration*. O banco de dados MySQL é utilizado como repositório de dados formando um *Data Warehouse* (DW) por onde os indicadores podem ser consultados. A ferramenta PowerBI é utilizada para geração dos dashboards para análise dos dados e geração de novos indicadores. Uma aplicação Web é apresentada visando catalogar os dados do COVID-19 e por meio da qual os dashboards são consultados. O trabalho apresenta ainda como resultados a discussão sobre os indicadores apresentados e trabalhos futuros sobre a implementações de novas tecnologias para melhorar a análise dos dados gerados.

Palavras-chave: Data Warehouse. *Business Intelligence*. CodeIgniter. COVID-19.

¹ Discente do Curso de Engenharia de Software do Centro Universitário Academia – UniAcademia. Endereço: Rua Senador Salgado Filho nº290, Bom Pastor, Juiz de Fora (MG). Celular: (32-988889597). E-mail: terra.saha@gmail.com

² Docente do Curso de Engenharia de Software do Centro Universitário Academia. Evaldo de Oliveira da Silva.

1 INTRODUÇÃO

Em dezembro de 2019, segundo o site Carreira Médica (2020)³, os primeiros casos do coronavírus (COVID-19) tiveram origem no mercado de frutos do mar da cidade de Wuhan, China. O COVID-19 foi identificado como uma doença viral que causava problemas respiratórios agudos. A incidência aumentou de forma exponencial e mais tarde se espalhou pelo mundo, sendo no dia 11 de março de 2020 declarada pela Organização Mundial da Saúde (OMS), como uma pandemia. Depois sendo declarada uma Emergência de Saúde Pública de Âmbito Internacional (PHEIC). Com a alto número de casos, os dados com números de casos e óbitos passaram a ser registrados diariamente e disponibilizados no site da OMS "<https://covid19.who.int/table>", no site do governo do Brasil "<https://covid.saude.gov.br/>", sites hospitalares, entre outros. Podendo em alguns desses sites, serem feitos o download dos arquivos com extensão (CSV).

Uma solução para analisar processos de saúde de forma integrada para alcançar metas e melhorias de processo é a utilização de *Business Intelligence* (BI). *BI* é um conjunto de metodologias, processos e tecnologias que são empregadas para coletar, integrar, analisar e disponibilizar dados transformando-os em informações significativas e úteis para permitir "insights estratégicos, táticos e operacionais" mais eficazes e tomada de decisão (SALIMON; MACEDO, 2017). Além disso, o BI da transparência de informações para o público.

Neste contexto, este trabalho visa o desenvolvido de um sistema web para armazenar os arquivos e realizar as etapas de Extração, Transformação e Carga (ETL) com o *Pentaho Data Integration*, alimentando um banco de dados MySQL, que servira como *Data Warehouse* (DW) e este por sua vez sendo a base de dados para a geração de dashboards no Power BI. Sendo estes dashboards disponibilizados na aplicação para auxiliar profissionais de saúde na análise dos números sobre os casos de COVID-19.

A aplicação web será desenvolvida a partir do framework CodeIgniter⁴ e utilizando um banco de dados relacional, Oracle⁵. Será usada uma base de dados fictícia, baseado no banco de dados já existente do hospital Santa Casa de

³ <https://www.sanarmed.com/coronavirus-origem-sinais-sintomas-achados-tratament>

⁴ <https://codeigniter.com/>

⁵ <https://www.oracle.com/br/index.html>

Misericórdia de Juiz de Fora⁶.

A partir da aplicação descrita e da análise dos dados extraídos, será possível desenvolver indicadores no BI e disponibiliza-los na aplicação para análise da evolução do número de casos de COVID-19 e para auxiliar na tomada de decisão do hospital, contendo indicadores de Juiz de Fora, Minas Gerais, região Sudeste, Brasil e de outros países.

Os indicadores propostos neste trabalho foram sugeridos através de uma pesquisa online, enviada para cerca de cinquenta profissionais e obtendo resposta de quarenta e quatro, sendo estes profissionais, médicos ou enfermeiros. Os indicadores sugeridos foram: número óbitos acumulados, número óbitos diários, número de casos acumulados, número de casos diários, número de recuperados e em acompanhamento, taxa de incidência, taxa de mortalidade, taxa de letalidade, casos por idade, casos por sexo, casos e óbitos por semana e por mês. Realizando uma análise comparativa desses indicadores entre regiões e estados do Brasil e os dados do Brasil com outras regiões no mundo. Também foi realizado um levantamento com indicadores com o número de leitos de enfermagem e UTI ocupados com COVID-19, a taxa de ocupação desses leitos e o número de leitos vagos na cidade de Juiz de Fora. Estes dados foram obtidos através de dados da Prefeitura de Juiz de Fora, contendo uma comparação com os números em cada hospital da cidade.

Os indicadores para análise propostos neste trabalho visam auxiliar em tomadas de decisão, não só dos hospitais da cidade em relação à infraestrutura e disponibilização de leitos para o isolamento dos pacientes com o vírus, mas também de pacientes na hora de escolher o hospital mais adequado para seu atendimento. Os dashboards⁷ também podem auxiliar hospitais, a população e o governo, ao indicar tendências do vírus no mundo e no país em relação a novas ondas e tomar medidas provisórias.

Este artigo segue estruturado em seções. A Seção 2 discute as soluções para o desenvolvimento de BI, onde é abordado os conceitos do framework CodeIgniter, de ETL, de *Data Warehouse* e de BI. A Seção 3 apresenta a metodologia do trabalho.

⁶ <https://santacasajf.org.br>

⁷ Dashboard é um termo criado por Stephen Few em 2013, onde ele usa a definição: “Um Dashboard é a apresentação visual das informações mais importantes e necessárias para alcançar um ou mais objetivos do negócio, consolidadas e ajustadas em uma única tela, para que a informação possa ser monitorada de forma ágil”. Few, Stephen – 2013. (Fonte: <http://www.perceptualedge.com>)

A Seção 4 apresenta os indicadores implementados por meio de uma aplicação Web e uma breve explicação sobre duas operações OLAP, o *Drill Up* e o *Drill Down*. Por fim, a Seção 5 faz as considerações finais deste trabalho, as contribuições e sugestões de trabalhos futuros.

2 SOLUÇÕES PARA O DESENVOLVIMENTO DE APLICAÇÕES DE BI

Neste capítulo serão apresentados os conceitos a serem contextualizadas neste trabalho. Sendo dividido em três etapas principais: o processo de desenvolvimento do sistema, as etapas de ETL e sobre o DW, e o Business Intelligence. No primeiro tópico, são apresentados o framework e o SGBD utilizado no desenvolvimento do sistema. Seguido pela especificação de ETL e da apresentação da ferramenta utilizada, além de uma explicação sobre o DW. Por fim, para apoiar o estudo, será conceituado e apresentado o Business Intelligence.

2.1 SISTEMA WEB

Para o desenvolvimento da aplicação proposta neste trabalho, foi necessária a utilização de uma aplicação web, para catalogar os arquivos de dados extraídos dos sites da OMS, do Governo, da Prefeitura de Juiz de Fora e de outros sites, para ter maior controle sobre os arquivos, as datas que estes dados foram baixados e o responsável pelo upload no sistema. A aplicação tem como finalidade também, disponibilizar o formulário para os médicos responder e os painéis de indicadores desenvolvidos no Power BI.

Foi adotado para o desenvolvimento do sistema, o framework CodeIgniter (CI). O CI é um kit de ferramentas para aplicações PHP (ANDRADE, 2011) e permite ser implementado por meio do padrão de projeto MVC (*Model-View-Controller*), e também pelo padrão HMVC (*Hierarchical Model-View-Controller*). O CI tem como principal objetivo desenvolver aplicações de forma mais rápida, mantendo o foco em minimizar a quantidade de códigos necessários para uma determinada tarefa (ANDRADE, 2011). Para o armazenamento dos dados deste sistema, foi utilizado o Oracle Database, um sistema gerencial de banco de dados, executado a partir de um programa PHP, através de comandos SQL.

2.2 EXTRACT, TRANSFORM, LOAD (ETL) E DATA WAREHOUSE (DW)

O ETL é um processo para extrair dados de um sistema de Banco de Dados (BD), sendo esses dados processados, modificados e posteriormente inseridos num outro BD. (Ferreira; João, Miranda; Miguel, Abelha; Antônio, Machado; José, 2010). Este outro BD é conhecido como *Data Warehouse* (DW). Segundo (CHAUDHURI; DAYAL, 1997) DW é uma coleção de tecnologias para suporte de decisão, destinadas a capacitar o trabalhador do conhecimento para tomada de decisões melhores e mais rápidas.

Antes dos dados serem carregados no DW, é realizado um processo de extração dos arquivos, sendo na etapa seguinte transformados e “limpos”, ou seja, são corrigidos, padronizados e tratados os desvios e inconsistências, transformando de acordo com as regras de negócios (ELIAS, 2014).

A utilização de um SGBD diferente para ser o repositório de dados, é para não sobrecarregar o banco de dados da aplicação e impedir que o sistema tenha uma sobrecarga e trave. (SARKIS, 2001) O DW apresenta duas características essenciais: ser integrado e histórico, proporcionando respectivamente bases para análise de tendências e observações por um espaço de tempo. Outra característica importante do DW, é não ser volátil, ou seja, os dados são carregados em blocos, servindo apenas para consultas e geração de relatórios para tomada de decisão.

Para o trabalho proposto, será utilizado o MySQL⁸, para ser o armazém de dados, sendo o mesmo *Open Source* e a ferramenta *Pentaho Data Integration*⁹ (PDI), para realização do ETL. Segundo (Arcenio, 2014) o PDI, permite aos usuários visualizar, manipular e analisar os dados em suas múltiplas dimensões do cubo OLAP.

OLAP é uma das ferramentas mais importantes do processo de ETL e do Business Intelligence. As ferramentas OLAP permitem a geração de relatórios, a análise de um grande volume de dados e a obtenção de informações estratégicas que podem facilitar a tomada de decisão (DIAS, 2014).

A consulta do DW é do tipo *On-Line Analytical Processing* (OLAP), enquanto da consulta de um Banco de Dados possui tipo de consulta *On-Line Transaction Processing* (OLTP). A diferença entre ambos, segundo (ANJOS; TOMIOKA, 2015) se dá pelo fato dos dados do DW, terem passado pelos processos de ETL. Enquanto o

⁸ <https://www.mysql.com/products/workbench/>

⁹ <https://www.hitachivantara.com/>

primeiro visa uma análise rápida para tomadas de decisão, o outro visa capacitar inserções, atualizações e remoção de dados, obedecendo regras de normalização.

A tabela 1 compara as aplicações OLTP e OLAP extraídas do trabalho de Neri (2002).

OLTP	OLAP
Controle operacional	Tomada de decisão
Atualização de dados	Análise de dados
Pequena complexidade das operações	Grande complexidade das operações
Não armazena dados históricos	Armazena dados históricos
Voltada ao pessoal operacional	Voltada aos gestores do negócio

Tabela 1 – Diferenças entre OLTP e OLAP (Neri, 2002)

2.3 BUSINESS INTELLIGENCE

Segundo [SERRA, 2019], define BI em seu livro “POWER BI - Do Tratamento de Dados a Criação de Painéis Inteligentes”, como um processo de conhecimento interno, ou seja, os dados devem ser coletados, tratados e organizados, para que possa ser feita uma análise e visibilidade através de painéis e Indicadores.

Como dito anteriormente, o BI coleta, integra, analisa e disponibiliza dados. Estes dados geram informações, que geram conhecimento. Este processo é de fundamental importância para escolher os indicadores ideais e criar dashboards precisos, podendo ser mais eficazes na tomada de decisão e auxiliando na transparência de dados para todos. Para realizar o desenvolvimento dos analytics, foi utilizado a ferramenta da Microsoft, Power BI. Sendo o Power BI Desktop para a geração dos dashboards e a opção web, no e-mail, para gerar um iframe para inserir os painéis gerados dentro do sistema web.

[SERRA, 2019] define em cinco passos, um fluxo de como deve ser o processo de desenvolvimento de um BI, iniciando pela coleta de dados, definindo bem, quais serão as fontes de dados para realizar a análise. O segundo passo, é realizado a organização e tratamento dos dados. No terceiro passo, realizar uma análise dos dados, para verificar se os dados disponíveis são suficientes para quantificar o estudo. A quarta etapa é a criação da visualização, ou seja, o desenvolvimento dos dashboards para visualizar os dados. Por último, planejar o futuro, o painel de

indicadores servirá para de base para tomada de decisão e no planejamento de ações futuras.

No caso deste trabalho, as fontes principais, serão extraídos dos sites da OMS e do Governo. O segundo passo, os dados são organizados e tratados, através do *Pentaho Data Integration*. No terceiro passo, foi realizado vários cálculos, utilizando formulas DAX¹⁰, para criar novos indicadores. No quarto passo, foi utilizado para o desenvolvimento do painel de indicadores, o Microsoft Power BI¹¹. Na última etapa do processo, é necessário profissionais especializados para a tomada de decisão dentro do hospital. Por isso, os painéis serão apresentados e validados por alguns médicos da Santa Casa.

3 METODOLOGIA

Como se trata de uma análise de indicadores na área da saúde, foi necessário buscar informações com médicos e enfermeiros. Foi realizado uma pesquisa online, para descreverem quais os principais indicadores sobre o corona vírus achavam mais importante. A partir das respostas, foi realizado um levantamento nos dados extraídos para saber quais indicadores seria possível criar.

Por ser analista de sistemas no hospital Santa Casa de Misericórdia de Juiz de Fora, foi realizado uma parceria com o hospital, ajudando no processo de pesquisa com os funcionários do hospital e da avaliação prévia dos dashboards gerados. É válido ressaltar que todas as ferramentas utilizadas, são open source e todas utilizadas pela TI do hospital, podendo ser implementado futuramente no sistema e no BI com os dados dos pacientes internados com COVID no hospital.

A parceria ajudou também com acesso ao banco de dados do hospital e com recursos como e-mail corporativo office 365, para a utilização dos formulários e do Power BI Web, e no conhecimento das regras de negócio impostas na interface. Como o objetivo do sistema desenvolvido no trabalho é possibilitar o armazenamento de arquivos de dados sobre o COVID-19, foi constatado que as entidades: setor, usuário, arquivo, tipo de arquivo, e que os recursos: de listagem de arquivos, listagem de usuários, listagem de setores, autenticação de usuário, recuperação de senha, cadastro de usuário, cadastro de setor, upload de arquivo,

¹⁰ DAX (*Data Analysis Expressions*) são expressões de fórmulas usadas no Power BI, incluindo funções, operações e valores para realizar cálculos avançados.

¹¹ [https:// powerbi.microsoft.com/pt-br/](https://powerbi.microsoft.com/pt-br/)

edição de setor, edição de usuário, edição de arquivo, edição do site do arquivo, exclusão de setor, exclusão de arquivo, download do arquivo, além da desativação ou ativação do usuário, seriam essenciais para o funcionamento do sistema.

Os arquivos no sistema são categorizados por site de onde foi realizado o download, e pela data que foi atualizado os dados do arquivo, permitindo o agrupamento dos arquivos futuramente e facilitando na localização do arquivo dentro do sistema facilmente. Para que o arquivo possa ser armazenado, é necessário que haja um usuário previamente autenticado. Esse usuário terá permissão de edição e exclusão apenas dos arquivos no qual fez o upload. Sendo disponível visualização para qualquer funcionário e apenas funcionários vinculados ao setor “Tecnologia da Informação” possuem permissão completa, de editar e excluir qualquer arquivos.

Dessa forma, foram definidos os requisitos para que o sistema proposto possa funcionar de forma a atender os padrões de gestão do hospital. Esses requisitos consistem em:

O sistema deve conter um mecanismo para autenticação do usuário; a aplicação deve permitir que apenas usuário vinculado ao setor “Tecnologia da Informação” visualize os dashboards de setor e de usuário no menu inicial, visualize a tela de setor e a tela de usuário, criar, editar e excluir setores, criar, editar e excluir funcionários; redefinir senha de outros funcionários; o sistema não deve permitir que um usuário seja excluído, apenas ativado ou desativado, por usuários vinculados ao setor “Tecnologia da Informação”; possibilitar que o usuário possa recuperar sua senha; a aplicação deve permitir que o usuário selecione o arquivo que será realizado o upload; o usuário logado deve ter permissão para realizar a edição e exclusão apenas de seus arquivos, caso não seja vinculado ao setor “Tecnologia da Informação”, caso seja, tem permissão irrestrita; o sistema deve ter uma tela que permita o usuário logado visualizar somente seus arquivos; deve permitir que o usuário logado possa visualizar o analytics desenvolvido com os dados do COVID-19 em Juiz de Fora, no Mundo e no Brasil; deve ser permitido que apenas médicos e enfermeiros possam visualizar a tela do formulário de pesquisa sobre os indicadores ideais sobre o vírus.

Os arquivos armazenados no sistema, sofreram um processo de Extração, Transformação e Carga (ETL) de dados, utilizando a ferramenta *Pentaho Data Integration*, onde serão armazenados os dados em um DW, sendo utilizado o

MySQL. Após este processo, será desenvolvido o Analytics, conforme os indicadores sugeridos pelos médicos através da pesquisa realizada, contendo informações mundiais, divulgadas pela OMS, nacionais, regionais e estaduais, divulgadas pela ANS (Agência Nacional da Saúde) e pelo governo, e municipais, liberadas pela Prefeitura de Juiz de Fora. Isso terá como objetivo, auxiliar na tomada de decisões sobre o COVID-19.

4 APLICAÇÃO DE BUSINESS INTELLIGENCE PARA ANÁLISE DE DADOS SOBRE O COVID-19

A aplicação web, terá como objetivo principal armazenar os arquivos do COVID-19, e disponibilizar os painéis de indicadores para todos que possuem acesso ao sistema, podendo analisar e acompanhar a evolução do vírus. Também será possível gerenciar setores e usuários.

O desenvolvimento do sistema, foi realizado de forma a cumprir os requisitos funcionais previamente levantados. O primeiro requisito do sistema exige uma autenticação de usuário no sistema. Com isso, foi desenvolvida uma tela de login para que o usuário possa informar suas credenciais de acesso, matrícula e senha, e se autenticar no sistema. O pedido de autenticação acontece através de uma requisição *HTTP* à interface de acesso sistema. Para acessar o sistema, o funcionário tem que estar cadastrado e ativo, caso não cumpra esses requisitos ou se as credenciais estejam incorretas, dará uma mensagem de erro e impedindo o acesso ao sistema. Caso as credencias informadas pelo usuário sejam válidas e o usuário esteja ativo, será verificado o setor do usuário logado, para realizar a segurança das informações e funcionalidades do sistema, após este processo o usuário será redirecionado à tela de inicio da aplicação.

A classe *Login* possui a responsabilidade de fazer o gerenciamento de login e logout no sistema. O método responsável por realizar a autenticação do usuário no sistema é o *Logar*, que recebe os parâmetros de entrada *login*, que é a *matricula* e *password* e realiza a requisição *HTTP* à interface do sistema utilizando o método *POST*.

Caso a autenticação tenha falhado, quer dizer que o usuário não está ativo ou suas credenciais estão incorretas ou não está cadastrado. Caso a verificação seja verdadeira, quer dizer que o usuário se autenticou com sucesso no sistema e

ele está ativo, sendo gravado em uma variável global, o setor do funcionário logado. O setor é parte fundamental no sistema, pois é de acordo com o setor, que são realizados os níveis de acesso de cada usuário. A Figura 1, ilustra a autenticação na classe M_Usuario, responsável pelo gerenciamento de usuários no sistema. Através do método AcessarSistema, será realizado uma busca no banco de dados, para conferir se o usuário em questão está digitando de forma corretas os dados e se está ativo.

FIGURA 1 — Autenticação de usuário.

```

public function AcessarSistema($login, $senha){
    $sql = "
    SELECT
        u.id_usuario,
        u.cd_matricula,
        u.nm_usuario,
        u.cpf_usuario,
        u.dt_nascimento,
        u.email_usuario,
        u.cel_usuario,
        u.endereco_usuario,
        s.nm_setor,
        u.ativo
    FROM
        usuario u
    INNER JOIN setor s on s.id_setor = u.id_setor
    where u.cd_matricula = '$login' AND u.psw_usuario = '$senha' and ativo='S'
    ";
    $stid = oci_parse($this->conexao, $sql);
    oci_execute($stid);
    $retorno = 0;
    $i=0;
    while (($row = oci_fetch_array($stid, OCI_ASSOC + OCI_RETURN_NULLS)) != false) {
        $i++;
        $retorno = array(
            'contador' => $i,
            'codigo' => $row['ID_USUARIO'],
            'matricula' => $row['CD_MATRICULA'],
            'nome' => $row['NM_USUARIO'],
            'cpf' => $row['CPF_USUARIO'],
            'data nascimento' => $row['DT_NASCIMENTO'],
            'email' => $row['EMAIL_USUARIO'],
            'celular' => $row['CEL_USUARIO'],
            'endereco' => $row['ENDERECO_USUARIO'],
            'setor' => $row['NM_SETOR'],
            'ativo' => $row['ATIVO']
        );
    }
    return $retorno;
}

```

Fonte: Autor 2020

Ao logar na aplicação um dashboard é apresentado com a quantidade de arquivos, de setores e funcionários cadastrados no sistema. Sendo apenas o dashboard, com a quantidade de arquivos disponíveis para todos os usuários. Os demais somente estarão liberados para funcionários da TI.

Ao logar, o sistema apresenta o nome do usuário logado, o setor deste funcionário, o dashboard e as opções do menu, também tendo controle por nível de acesso dependendo do setor do funcionário logado, sendo visíveis para todos apenas as telas de analytics¹² e os arquivos do COVID-19. A tela de pesquisa é

¹² Analytics é o mesmo que Painel de Indicadores.

acessada somente por médicos e enfermeiros, onde responderão quais os indicadores sobre o vírus acham mais importante para auxiliar no desenvolvimento dos *Analytics*. Também é possível o usuário logado, ver suas próprias informações e reconfigurar sua senha de acesso ao sistema.

O segundo requisito funcional estipulado para o aplicativo é de disponibilizar uma tela de cadastro de setor. Essa tela é acessível apenas para usuários que sejam do setor "Tecnologia da Informação", a partir da tela inicial, no menu e disponibiliza ao usuário uma tabela com todos os setores já cadastrados, um botão para excluir o setor e botões para adicionar ou editar o setor, sendo um modal, contendo um campo para que possa ser informado o nome do setor.

O terceiro requisito funcional estipulado, é a o gerenciamento do usuário. Onde é possível adicionar novo funcionário, edita-lo, podendo inativar o funcionário ao invés de excluí-lo, possibilitando a manutenção dos dados de qual funcionário fez o upload dos arquivos, e um botão para reset da senha.

O quarto requisito, é o gerenciamento dos arquivos do COVID-19, onde tem duas telas para realizar esse gerenciamento. A primeira é a tela de lista de arquivos, onde todos podem visualizar a lista de arquivos e baixa-los, sendo somente quem for do setor "Tecnologia da Informação", autorizado para excluir ou editar estes arquivos. A segunda tela, é a tela dos arquivos do usuário logado, onde é possível editar e excluir seus próprios arquivos.

Ao adicionar novo arquivo, é necessário informar a data que foi atualizado os dados do arquivo, o site de onde foi realizado o download e escolher o arquivo (Figura 2), sendo calculado pelo sistema a data de cadastro do arquivo no sistema.

FIGURA 2 — Tela de gerenciamento de Arquivos – Adicionar Arquivo.

Adicionar arquivo de COVID-19

dd/mm/aaaa

Escolher arquivo Nenhum arquivo selecionado

http:// Digite o endereço da fonte

Fechar

Fonte: Autor 2020

Para implementação do sistema, foi adicionado um iframe (Figura 3), uma tag

que permite incluir um site com formato menor dentro de um site principal. Foi utilizado um formulário do office 365¹³, na tela de pesquisa, para que médicos e enfermeiros possam responder e sugerir indicadores sobre o COVID-19. Também foi utilizado o iframe para disponibilização dos painéis de indicadores, desenvolvidos pelo Microsoft Power BI Desktop e publicado no Power BI Web¹⁴, também disponibilizado pelo office 365.

FIGURA 3 — Iframe para inclusão do Analytics do Power BI.

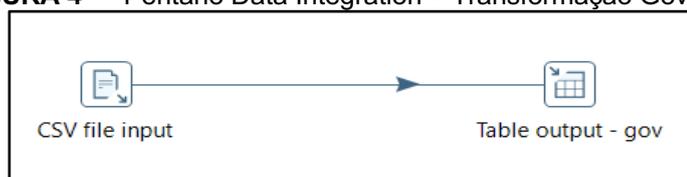
```
<iframe width="1500" height="900"
src="https://app.powerbi.com/view?r=eyJrIjoiYzgyN2
QxNzUtNWMyYy00MGViLTk0MmYtZGIzN2Y2NmFkMzEzIiwidCI6
ImRhZjBiNzk3LTFjYjYtNGMzOC04MzY3LWY3MzUyMTcxYWE3OS
J9&pageName=ReportSectionc7993374c1ee7825d91a"
frameborder="0" allowFullScreen="true"></iframe>
```

Fonte: Autor 2020

Após os requisitos principais que compõem o sistema ficarem prontos, foi realizado as etapas de ETL. Para a realização das etapas de ETL, foi adotado o suíte *Pentaho Data Integration* (PDI). O PDI, possibilita trabalhar com vários tipos de base de dados, como exemplo um SGBD, planilhas, arquivos de texto, APIs, entre outros. Para o projeto em questão, foram utilizados arquivos do tipo CSV, passando pelas etapas de extração de dados, transformação e carga em um DW MySQL, que assim como o Oracle, é um SGBD relacional, que será o armazém dos dados.

Foram utilizados dois métodos de transformação. O primeiro com os dados do governo, sendo a tabela truncada, ou seja, toda vez que for executada a transformação, todos os dados serão apagados e inseridos novamente, junto com os novos registros. O Pentaho permite a configuração para realizar a carga dos dados no DW, no caso é usado a base de dados no MySQL, covid_dw e a tabela governo, além de configurar se será truncado ou não os dados. A Figura 4, apresenta as etapas de ETL realizadas nos dados do Governo.

FIGURA 4 — Pentaho Data Integration – Transformação Governo.



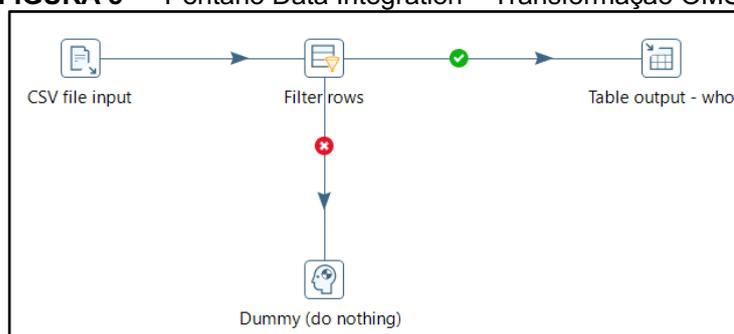
Fonte: Autor 2020

¹³ <https://forms.office.com/>

¹⁴ [https:// app.powerbi.com/](https://app.powerbi.com/)

O segundo método foi utilizando os dados da OMS (Figura 5), onde é utilizado um mecanismo de filtro (Filter rows), filtrando por data. Quando a data for menor que a última atualização dos dados, ele joga para um step nomeado Dummy, que servirá como um local para armazenar os valores “errados”, enquanto que os registros com a data filtrada, serão inseridos no DW. A prática do uso de filtros, pode auxiliar na velocidade da execução da transformação e carga dos dados no DW, dependendo da quantidade de dados que será realizado o processo de ETL.

FIGURA 5 — Pentaho Data Integration – Transformação OMS.

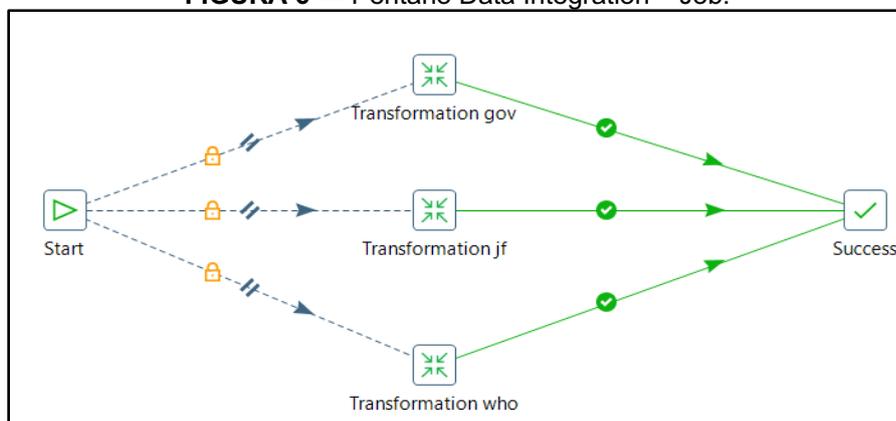


Fonte: Autor 2020

Para implementação do projeto, também foi criado um job (Figura 6), para poder executar as 3 transformações criadas, tabela com dados do Governo, tabela com dados da OMS e tabela com dados de Juiz de Fora, ao mesmo tempo.

Foi utilizado hops (linha que liga um step a outro) executando em paralelo e incondicional, ou seja, todas as transformações irão iniciar ao mesmo tempo e independente se der erro ou não na execução de um dos steps, os outros podem ser realizados e finalizados com sucesso.

FIGURA 6 — Pentaho Data Integration – Job.



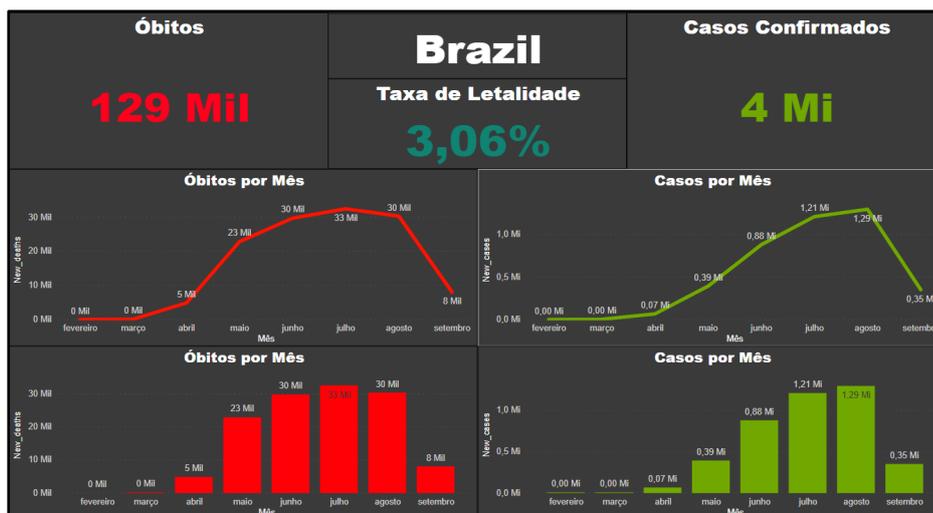
Fonte: Autor 2020

Outra utilidade do job, é a possibilidade de programar a execução dele, de preferência estando em um servidor, pode ser programado para o job rodar por um determinado intervalo de tempo, podendo ser um intervalo em segundos, minutos dias, semanas ou meses. Além de possibilitar configurar a hora exata, ou o dia da semana, dia do mês que será sempre executado o job.

Por fim, após ter os dados armazenados no DW, foi possível desenvolver os analytics usando a conexão com o MySQL. Para a geração dos dashboards, foi analisado os dados disponíveis e os indicadores sugeridos por especialistas da área da saúde, como por exemplo, número de óbitos, número de casos, taxa de incidência, taxa de letalidade, taxa de mortalidade, mapeamento da capacidade dos leitos dos hospitais de Juiz de Fora. Para demonstrar alguns dos indicadores, foi escolhido a Figura 7, com o painel com alguns dados do Brasil, retirados do site da OMS. Nele é representado a quantidade de óbitos totais, o número de casos confirmados, a taxa de letalidade e alguns gráficos que mostram a curva e evolução do número de óbitos e casos mês a mês.

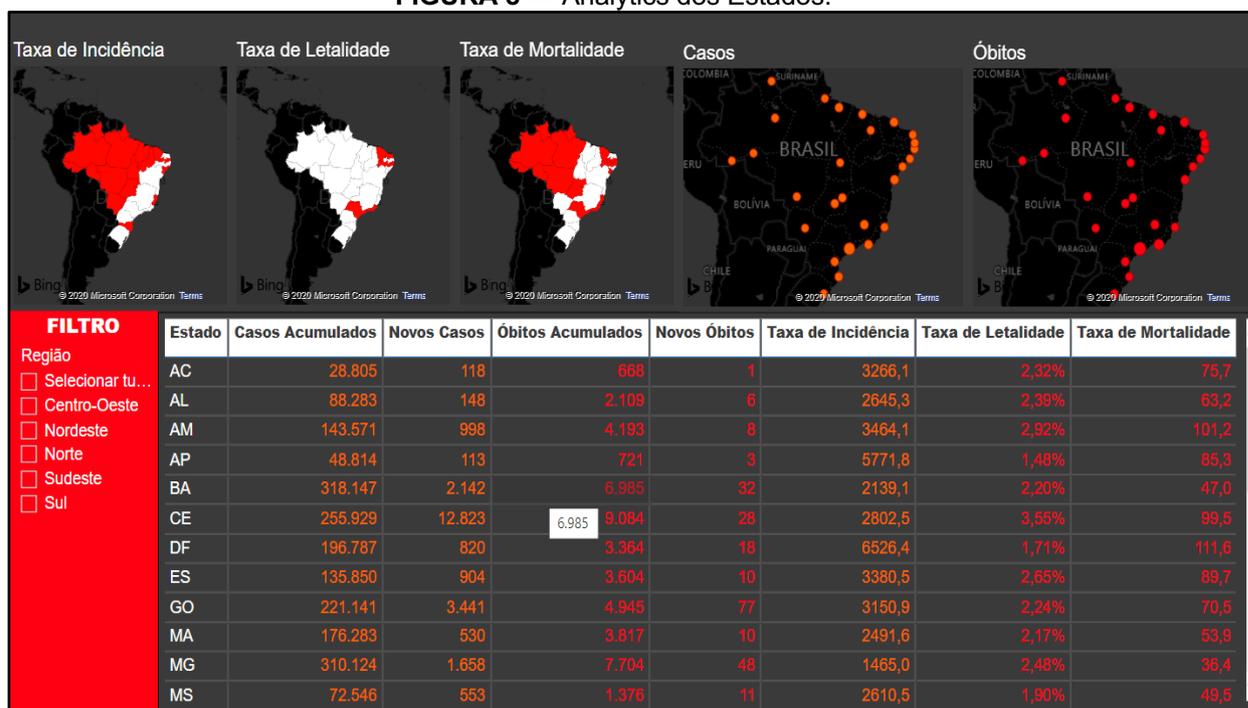
O Power Bi possui uma série de designs para compor o analytics, uma delas é a possibilidade de colocar um mapa e como no exemplo abaixo (Figura 8), mostrando vários indicadores, como taxa de letalidade, taxa de incidência, taxa de mortalidade, número de casos e de óbitos em cada estado, além da utilização de uma personalização condicional, onde foi utilizado como condição as taxa de letalidade, incidência e mortalidade no Brasil, e comparando em cada estado. Onde as taxas forem menores que as taxas no país, fica com uma marcação na cor branca e onde as taxas forem maiores, a marcação fica em vermelho.

FIGURA 7 — Analytics Brasil.



Fonte: Autor 2020

FIGURA 8 — Analytics dos Estados.



Fonte: Autor 2020

Foram utilizado neste trabalho, dois operadores das ferramentas OLAP, o *Drill Down* e o *Drill Up*. Um é o oposto do outro, enquanto o primeiro aumenta o nível de detalhamento dos dados, ou seja, diminui a granularidade, o segundo diminui o nível de detalhamento, aumenta a granularidade. De acordo com o site Devmedia¹⁵, a granularidade determina quais tipos de consulta podem ser feitas no DW.

¹⁵ <https://www.devmedia.com.br/um-estudo-sobre-as-ferramentas-olap/6691>

Para uma análise mais detalhada dos números no país, foi desenvolvido uma matriz, dividindo os dados por região e por estado. Onde é possível realizar o operador *Drill Up* (Figura 9) para ter uma amostra menos detalhada, apenas por regiões, e o operador *Drill Down* (Figura 10), apresentando os valores dos estados de determinada região

FIGURA 9 — Drill Up.

Regiões	Casos	Óbitos	Taxa de Letalidade	Taxa de Mortalidade	Taxa de Incidência
Centro-Oeste	618.545	13.198	2,13%	81,0	3795,4
Nordeste	1.361.267	39.865	2,93%	69,9	2385,2
Norte	644.125	15.216	2,36%	82,6	3494,8
Sudeste	1.732.228	66.681	3,85%	75,5	1960,2
Sul	612.976	12.534	2,04%	41,8	2044,9
Total	4.969.141	147.494	2,97%	70,2	2364,6

Fonte: Autor 2020

FIGURA 10 — Drill Down.

Regiões	Casos	Óbitos	Taxa de Letalidade	Taxa de Mortalidade	Taxa de Incidência
Centro-Oeste	618.545	13.198	2,13%	81,0	3795,4
Nordeste	1.361.267	39.865	2,93%	69,9	2385,2
Norte	644.125	15.216	2,36%	82,6	3494,8
Sudeste	1.732.228	66.681	3,85%	75,5	1960,2
ES	135.850	3.604	2,65%	89,7	3380,5
MG	310.124	7.704	2,48%	36,4	1465,0
RJ	275.415	18.883	6,86%	109,4	1595,2
SP	1.010.839	36.490	3,61%	79,5	2201,4
Sul	612.976	12.534	2,04%	41,8	2044,9
Total	4.969.141	147.494	2,97%	70,2	2364,6

Fonte: Autor 2020

O código fonte completo do sistema desenvolvido está disponível em repositório do *github*¹⁶ e os painéis de indicadores estarão disponíveis na web também através dos links disponibilizados no final do artigo.

5 CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS

Considerando que atualmente os conceitos de *Business Intelligence*, *Big Data*, *Ciência de Dados*, estão em alta, existem vários segmentos da sociedade onde podem ser aplicadas para auxílio na tomada de decisão. Para o projeto foi definido

¹⁶ <https://github.com/Bene31/COVID-19>

usar esses conceitos na área da saúde, devida a situação mundial atual, com o COVID-19 e pela parceria com o hospital Santa Casa.

As dificuldades encontradas no trabalho proposto foram relacionadas às bases de dados disponibilizadas na internet. Esses dados sofreram alterações ao decorrer do desenvolvimento deste trabalho, como por exemplo, remoção ou adição de campos e alguns até sendo renomeados. Outra dificuldade, foi em relação à falta de alguns indicadores sugeridos na pesquisa.

O auxílio dos profissionais da saúde, possibilitou uma análise melhor para a obtenção dos indicadores utilizados neste trabalho e validados pelos mesmos, resultando numa tomada de decisão mais assertiva nas medidas a serem tomadas. Como por exemplo na distribuição de pacientes, leitos e dos médicos nos hospitais.

Além disso, o *Business Intelligence* é importante para reconhecimento da situação da pandemia e na possibilidade de transparência da situação municipal, estadual, nacional e mundial, permitem uma melhor conscientização da população, uma vez que a curva do número de casos na cidade pode subir com o relaxamento das pessoas em relação aos métodos de prevenção, distanciamento social e por causa de aglomerações e que muito se fala em novas ondas do vírus e de sua mutabilidade.

Hoje em dia, tudo e todos estão conectados à internet. A facilidade de acesso a informações na internet é um grande aliado para hospitais, já que a área da saúde cada vez mais está se tornando digital e tendo como foco os dados e sendo sua análise essencial para obter melhores resultados e melhor qualidade de atendimento. Isso cria inúmeras possibilidades de no futuro, ampliar este projeto utilizando dados de outras doenças e gerando novos indicadores de comparação entre elas.

ABSTRACT

This article consists in the construction of a Business Intelligence application to allow an analysis of indicators in the health area, based on data on the COVID-19 virus. Data were extracted from Ministry of Health and WHO (World Health Organization) websites. The indicators were assigned by health professionals based on the need to obtain useful information on the evolution of the number of cases of COVID-19 in the region of Juiz de Fora and other regions. For storing the data generated for the indicators, this work presents the steps of the ETL (Extract, Transform and Load) implemented through the Pentaho Data Integration tool. The MySQL database is used as a data repository forming a Data Warehouse (DW) where the indicators can be consulted. The PowerBI tool is used to generate dashboards for data analysis and generation of new indicators. A web application is manufactured to catalog the COVID-

19 data and through which the dashboards are consulted. The work also presents as results of the discussion on the indicators and future work on the implementation of new technologies to improve the analysis of the generated data.

Keywords: Data Warehouse. *Business Intelligence*. Codelgniter. COVID-19.

REFERÊNCIAS

ANDRADE, Fernando francisco de. **Desenvolvimento de aplicações web com a utilização dos frameworks Codelgniter e Doctrine**. 2011, 59 p. Dissertação (Trabalho de Diplomação em Sistema de Informação). Universidade Tecnológica Federal do Paraná, 2011. Disponível em: <<http://repositorio.roca.utfpr.edu.br/jspui/handle/1/611>> Acesso em: 17 set 2020

ANJOS, Matheus; TOMIOKA, Pedro. **Utilização de Data Warehousing Para Aferição Da Qualidade Interna de Software Apoiado Por Uma Abordagem de Interpretação de Métricas de Código Fonte - Um Estudo de Caso No Tribunal de Contas de União**. Brasília - DF, 2015, 220p. Disponível em: <<fga.unb.br/articles/0001/0524/tcc.pdf>> Acesso em: 04 nov. 2020.

ARCENIO, Luiz Fernando Stopa. **Integração dos Data Warehouse do AGHU das filiais MEC/EBSERH com Pentaho Data Integration**. 2014, 10 p. Setor de Informática do HUGD, Empresa Brasileira de Serviços Hospitalares (EBSERH), Dourados, MS. Disponível em: <<http://periodicos.unesc.net/sulcomp/article/view/1813/1715>> Acesso em: 30 out. 2020.

Carreira Médica. **Coronavírus (COVID-19): origem, sinais, sintomas, achados, tratamentos e mais**. 2020. Disponível em: <<https://www.sanarmed.com/coronavirus-origem-sinais-sintomas-achados-tratamentos>> Acesso em: 10 set 2020

CHAUDHURI, S.; DAYAL, U. **An overview of data warehousing and olap technology**. 1997, 10p. Disponível em: <<https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/sigrecord.pdf>> Acesso em: 04 nov. 2020.

DIAS, Franciele de Oliveira. **FERRAMENTAS OLAP: UM ESTUDO BASEADO NO SISTEMA PENTAHO**. 2014. 55 p. Trabalho de conclusão de curso (Graduação) - Instituto Municipal de Ensino Superior de Assis - IMESA, Assis, 2014. Disponível em: <<https://cepein.femanet.com.br/BDigital/arqTccs/1111320138.pdf>>. Acesso em: 17 out. 2020.

ELIAS, Diego. **Entendendo o Processo de ETL**. Disponível em: <<https://canaltech.com.br/business-intelligence/entendendo-o-processo-de-etl-22850/>> Acesso em: 18 set 2020

NERI, H. R. **Análise, Projeto e Implementação de um Esquema MOLAP de Data Warehouse utilizando SGBD-OR Oracle 8.1**. Universidade Federal da Paraíba - UFPB, 2002, 128p. Disponível em: <<http://dspace.sti.ufcg.edu.br:8080/xmlui/handle/riufcg/12111>> Acesso em: 04 nov. 2020.

SALIMON; Cláudia Cristina, MACEDO; Mary Caroline Skelton. **Aplicações de Business Intelligence na Saúde: Revisão de Literatura**. Journal of Health Informatics, v. 9, n. 1, 5 p, São Paulo, 2017. Disponível em: <<http://www.jhi-sbis.saude.ws/ojs-jhi/index.php/jhi-sbis/article/view/452/303>> Acesso em: 17 set 2020

SARKIS, Laura Costa. **DATA WAREHOUSE: O PROCESSO DE MIGRAÇÃO DE DADOS**. 2001, 142 p. Dissertação (Mestrado em Ciência da Computação). Universidade Federal de Santa Catarina, 2001. Disponível em: <<https://repositorio.ufsc.br/xmlui/bitstream/handle/123456789/80047/227423.pdf?sequence=1&isAllowed=y>> Acesso em: 17 set 2020

SERRA, Luis Gustavo N. **POWER BI - Do Tratamento de Dados a Criação de Painéis Inteligentes**. [S. l.]: Instituto Alpha Educação à distância e Editora, 2019. 270 p. ISBN 978-85-66018-37-x. *E-book* (34 p.).

APÊNDICE - LINKS DOS PAINÉIS DE INDICADORES DESENVOLVIDOS

Indicadores no Mundo

<https://app.powerbi.com/view?r=eyJrljoiNDk4OTU2ZDMtY2FiZS00NGRmLWE2MTQtNjU2ZTM5Y2MwOTIzIiwidCI6ImRhZjBiNzk3LTFjYjYtNGMzOC04MzY3LWY3MzUyMTcxYWE3OSJ9&pageName=ReportSection1f44a6bb674c586b7f06>

Indicadores no Brasil

<https://app.powerbi.com/view?r=eyJrljoiYzgyN2QxNzUtNWMyYy00MGViLTk0MmYtZGIzN2Y2NmFkMzEzIiwidCI6ImRhZjBiNzk3LTFjYjYtNGMzOC04MzY3LWY3MzUyMTcxYWE3OSJ9&pageName=ReportSectionc7993374c1ee7825d91a>

Indicadores em Juiz de Fora

<https://app.powerbi.com/view?r=eyJrljoiYmNiYTlxZTctNTQxZS00NGQ5LWE1NDItYmUyM2M4YmFhYTlkIiwidCI6ImRhZjBiNzk3LTFjYjYtNGMzOC04MzY3LWY3MzUyMTcxYWE3OSJ9&pageName=ReportSectionccf0e7ee9c8a6c2a8322>