

# **Desenvolvimento de um processo de ETL e uso de *dashboard* para Visualização de Casos de Covid**

**Elizabeth Santos, Edgar Noda**

<sup>1</sup> Instituto Federal de Educação, Ciência e Tecnologia - Campus Hortolândia  
(IFSP)  
CEP 13183-250 – Hortolândia – SP – Brazil

elizasantos22@gmail.com, edgar.noda@ifsp.edu.br

**Abstract.** *The scale of the COVID 19 pandemic and its impact on the world has resulted in a huge volume of data being generated and an even greater interest in its availability. However, it is common for this data to be distributed in different places and in different ways. Another factor to be considered is the speed at which these data are updated, this characteristic also increases the degree of complexity of data collection and presentation. Thinking about how to offer consolidated access, facilitating not only the process of collecting information, but also its subsequent analysis, this work proposes to create an ETL proposal on the data of the Covid 19 pandemic and presents them in the format of a dashboard and with access via WEB.*

**Resumo.** *A dimensão da pandemia de COVID 19 e seu impacto no mundo resultou em um enorme volume de dados sendo gerado e um interesse ainda maior na sua disponibilização. Entretanto, é comum que esses dados estejam distribuídos em locais e de formas diferentes. Outro fator a ser considerado é a velocidade que estes dados são atualizados, tal característica também aumenta o grau de complexidade da coleta e apresentação dos dados. Pensando em como oferecer um acesso consolidado, facilitando não apenas o processo de coleta das informações, mas também sua análise posterior, este trabalho se propõe a criar uma proposta de ETL sobre os dados da pandemia de Covid 19 e apresentá-los no formato de um dashboard e com acesso via WEB.*

## 1. Introdução

Vivemos em uma era de *Big Data*, os dados são gerados o tempo todo em uma velocidade, variedade e volume muito grande, e conseguir lidar com esses dados de forma eficiente é uma das tarefas mais importantes da área de dados. Os engenheiros de dados capturam dados de diferentes fontes, transformam esses dados e os deixam disponíveis através de um processo chamado *Datawarehouse*. Com os dados devidamente padronizados, os analistas e cientistas de dados podem realizar o seu trabalho de forma mais eficaz.

Com a pandemia do novo coronavírus o número de estudos sobre o assunto aumentou drasticamente e com eles o volume de dados produzidos e distribuídos. A explosão de dados aliado ao avanço no processamento possibilitou o crescimento da Ciência de Dados como base para auxiliar à tomada de decisões. Entretanto, é comum que esses dados não sejam apresentados de forma uniforme, estejam distribuídas em lugares diferentes, e na maioria das vezes não são estáticas, fatores estes que dificultam a coleta e apresentação dos dados. Segundo David Loshin (2021). Afirma que “As organizações mais bem-sucedidas terão uma estratégia clara e precisa em funcionamento, que reconheça data *integration* como pilar fundamental de sua vantagem competitiva”.

Pensando em como melhorar a análise e coleta dos dados referente a pandemia, este trabalho se propõe a criar um processo de ETL (Extração, transformação e carregamento dos dados) utilizando o *framework* de *Datawarehouse* conhecido o *BigQuery* ETZ (2020).

Com essa ferramenta se espera realizar uma busca diversificada dos dados, fazendo a extração desses dados brutos sendo eles estruturados ou não estruturados, após a extração guardar esses dados em uma plataforma de armazenamento na nuvem Google Cloud (2007). Realizar o tratamento e a organização dos dados, criar métricas e variáveis desses dados, realizar a unificação dessas informações em tabelas no *BigQuery*. Para a visualização dos dados, a utilização da plataforma da *Google Data Studio* Cetax (2022). Para apresentar os dados tratados e unificados em um *dashboard*.

Este trabalho aborda problemas atuais, tanto do ponto de vista de tecnologia, como de objeto de aplicação. As tecnologias a serem utilizadas representam uma grande tendência do mercado em como lidar com o crescente volume de dados sendo gerados e disponibilizá-los para os usuários finais. Do ponto de vista de aplicação, existe ainda a necessidade de ampliar a difusão de informações para a população a respeito da pandemia da covid. Nesse sentido, primeiro foi feito um levantamento junto a profissionais da área da saúde a respeito do que poderia ser feito a respeito dessa questão, tanto referente a doença, como relação a quais seriam as principais dúvidas que observavam na população. Após a análise deste levantamento, um ponto comum foi o problema no que diz respeito a disponibilização de dados e mais especificamente referente a coleta e tratamento de dados descentralizados da pandemia. A proposta busca então a disponibilização de uma forma unificada e clara para população em geral, no caso utilizando a plataforma *BigQuery* para centralizar os dados processados e apresentando-os em um *dashboard* na plataforma *Data Studio*. Assim, a abordagem procura disponibilizar de maneira fácil, rápida, visual e unificada as informações e de alguma forma, contribuir para o aumento da difusão das informações referentes a pandemia junto à população.

## 2. Fundamentação Teórica

Nesta seção abordará os principais conceitos que fundamentam os aspectos técnicos e o embasamento utilizado para o planejamento e desenvolvimento do processo de ETL proposto.

### 2.1. Big Data, Ciência e Engenharia de Dados

Como volume, velocidade e variedade são conceitos que aparecem com frequência nas definições de *Big Data* que aparecem na literatura, é importante entender como a literatura os descreve. Uma característica importante dos conjuntos de dados considerados como *Big Data* é o grande volume de dados RUSSOM (2011). Esse volume é consequência da grande quantidade de dados registrados na atualidade. Segundo McAfee & Brynjolfsson (2012), cerca de 2,5 EB (exabytes equivalem a um bilhão de gigabytes) são criados diariamente e ele projeta que esse número deve dobrar.

A ciência de dados combina os campos da estatística, método científico, inteligência artificial (IA) e análise de dados para extrair valor dos dados, as pessoas que trabalham em ciência de dados são chamadas de cientistas de dados e combinam uma variedade de habilidades para analisar dados coletados da Web, smartphones, clientes, sensores e outras fontes para obter insights acionáveis. A ciência de dados inclui a preparação de dados para análise, incluindo limpeza, agregação e manipulação de dados para realizar análises avançadas de dados.

A engenharia de dados é o campo que lida com a transformação dos dados brutos, esta é a primeira etapa do processamento de dados. Uma série de atividades que visam cumprir etapas, sendo elas o processo de coleta, armazenamento e distribuição de dados sob engenharia de dados, dar uso prático à vasta quantidade de informações disponíveis, usando um exemplo básico, digamos que você está organizando uma festa de casamento e tem uma lista de convidados em mãos, a lista não está em ordem, mas ao lado de cada convidado há informações sobre sua ligação com os noivos, família, colegas, vizinhos, amigos de infância. As informações da lista são brutas e uma boa maneira de aproveitá-las é dividir os convidados em grupos, um para a família, outro para os colegas de trabalho e assim por diante, transformando os dados originais.

Contudo, quando falamos de engenharia de dados, o trabalho será feito por software e algoritmos, dessa forma, essas tarefas envolvem muito conhecimento técnico, projeto de soluções a partir de bancos de dados, e conhecimento estratégico para alinhar as soluções aos objetivos da empresa ou do cliente, portanto, não é exagero dizer que, na analogia da construção civil, os profissionais de engenharia de dados são engenheiros e arquitetos ao mesmo tempo.

## 2.2. Processo de ETL

ETL é o processo de extrair, transformar e carregar dados. Surge como uma estratégia de análise e utilização de informações armazenadas em bancos de dados, sendo complexo ou simples. Por meio deste processamento é possível definir a forma como serão manipulados e a qualidade dos dados, assim transformando em dados confiáveis, etapas necessárias para se criar um ETL (Figura 1).

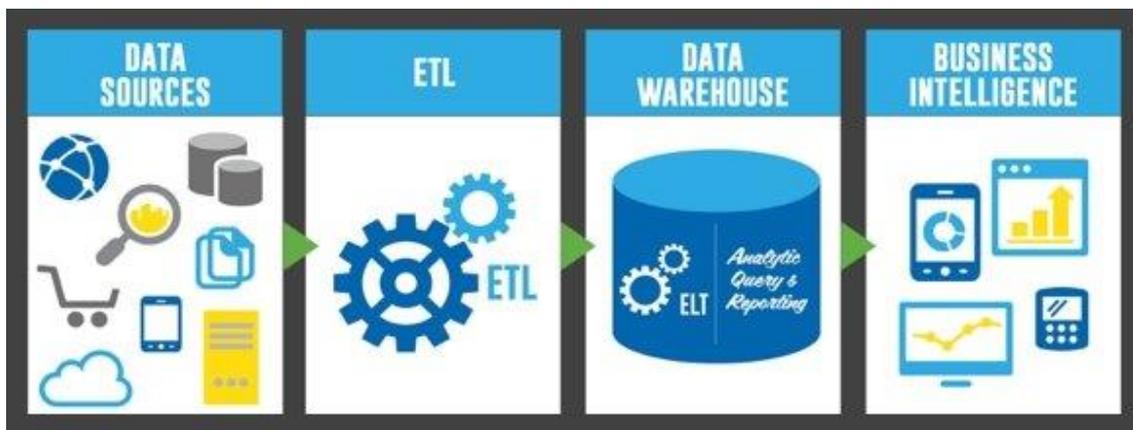


Figura 1. O Antes, durante e depois de um processo de ETL  
Fonte: <https://etz.com.br/entenda-o-processo-de-etl/>

Neste exemplo, podemos considerar que a primeira fase do processo seria a extração de dados em SQL. Neste processo de extração, os dados são convertidos e organizados em um formato único, assim tornando possível a manipulação nas próximas etapas. Como os dados são muito diferentes é essencial fazer a padronização desses dados.

Na fase de transformação, ocorre a adaptação das informações que foram analisadas e padronizadas na fase da extração, também cria-se nessa etapa os filtros para unificar informações como idade, localização, ou qualquer outro que possa ser útil para realizar futuras análises.

No último e terceiro passo do processo ETL, é necessário realizar o carregamento dos dados já organizados em um repositório, para essa fase duplicasse a tabela com as informações tratadas, assim realizando os ajustes necessários para corrigir novos desvios de fluxo mantendo um modelo de dados organizados. Pode-se criar um mapeamento de todos os padrões, tornando-os sempre acessíveis para utilizações futuras, esse processo não é necessariamente executado em um único ambiente de tratamento pode-se utilizar diversas aplicações para o processo, seja em nuvem ou não.

## 2.3. Linguagem de Modelagem Unificada UML

A UML (*Unified Modeling Language*) é uma linguagem que define uma série de artefatos que auxiliam na tarefa de representar, através de documentos, os sistemas orientados a objetos ao longo do seu processo de desenvolvimento, mas modelagem unificada não nos indica como desenvolver o software, como descreve Medeiros (2004), ela indica apenas as formas que podem ser utilizadas para representar um software em diversos estágios de desenvolvimento.

## 2.4. Cloud Computing

*Cloud computing* ou computação na nuvem, é uma tecnologia que permite acesso remoto a *softwares*, armazenamento de arquivos e processamento de dados por meio da *internet*, é uma alternativa para você acessar dados importantes de qualquer computador, em qualquer lugar. O benefício dessa abordagem não é apenas a redução dos custos com servidores locais, a *cloud* é uma maneira dinâmica de se ajustar às necessidades de recursos de infla estrutura ao seu negócio de maneira simples, segura e econômica.

## 2.5. BigQuery

O *BigQuery* é um serviço de armazenamento de dados de baixo custo e totalmente gerenciado do Google para análises com escala em petabytes, essa ferramenta é *NoOps*, ou seja, ela não requer uma infraestrutura responsável pelo gerenciamento nem um administrador de banco de dados. Com toda a infraestrutura oferecida pelo Google, o processo de coleta de dados possui uma performance excelente, desta forma, pode manter seus resultados atualizados de forma rápida, realizando consultas ou desenvolvendo relatórios em tempo real.

## 2.6. SQL

SQL significa “Structured Query Language”, ou “Linguagem de Consulta Estruturada”, em português, resumidamente, é uma linguagem de programação para lidar com banco de dados relacional (baseado em tabelas), criado para que vários desenvolvedores pudessem acessar e modificar dados de uma empresa simultaneamente, de maneira descomplicada e unificada. O SQL pode ser usada para analisar ou executar tarefas em tabelas, principalmente através dos seguintes comandos: inserir (*insert*), pesquisar (*search*), atualizar (*update*) e excluir (*delete*), também pode ser usada para realizar tarefas complexas, como por exemplo, escrever queries e fazer consultas ou manipulações.

## 2.7. Data Studio

*Google Data Studio* é uma ferramenta de business *analytics* gratuita bastante útil no marketing digital. É possível criar *dashboards* personalizados e atraentes para analisar dados e apresentar resultados de estratégias, seus dados em relatórios e painéis informativos são fáceis de ler e de compartilhar e totalmente personalizáveis. Por meio da visualização de dados, o *Google Data Studio* reúne métricas e indicadores que embasam a tomada de decisões e o direcionamento das estratégias, de forma que as empresas corram menos riscos e atinjam seus objetivos.

## 2.8. Google Forms

O *Google Forms* é um serviço gratuito para criar formulários online. Nele, o usuário pode produzir pesquisas de múltipla escolha, fazer questões discursivas, solicitar avaliações

em escala numérica, entre outras opções. A ferramenta é ideal para quem precisa solicitar feedback sobre algo, organizar inscrições para eventos, convites ou pedir avaliações. O *Google Forms* tem compatibilidade ampla, incluindo navegadores como *Firefox*, *Microsoft Edge* e, é claro, *Chrome*. Para iniciar, é preciso ter uma conta *Google*.

## 2.9. StarUML

StarUML, é um modelador de diagramas de *software* livre. Ferramenta útil, simples, concebida conjuntamente com o *Windows*, oferecendo por esta razão uma funcionalidade prática e viável a todos no setor de desenhos de fluxograma disponibilizando: uma caixa de diálogo, uma fácil manipulação do teclado e uma perfeita visualização geral do diagrama.

## 3. Trabalhos Correlatos

O processo ETL existe em qualquer tarefa de processamento de dados e é considerado um estágio chave de uma estratégia de utilização de dados, esse processo é responsável por garantir a qualidade da transformação de dados brutos em informações relevantes. O ETL é essencial para empresas que buscam consolidar dados em um ambiente de integração, no caso um data warehouse. Ele é um repositório central de dados que auxilia no processo de análise e tomada de decisões. Eles servem como *hubs* para consolidar dados de vários sistemas de origem e armazenar o histórico existente para análise e tomada de decisões de negócios. O processo de ETL pode ser aplicado a bancos de dados mais simples, como SQL, e pode até ser aplicado a mais sistemas complexos, como uma nuvem de *big data*.

Uma grande empresa que utiliza esse processo é a Dotz (2007) tendo como objetivo o tratamento de dados brutos em informações essenciais. A empresa Dotz trabalhou com diversas ferramentas da plataforma. “Utilizamos o *Cloud Composer* como orquestrador dos ETLs, cuja ingestão se dá por meio do *Dataflow*. O *BigQuery* foi usado como *Data warehouse*. Tudo foi automatizado utilizando o *Cloud Datastore* como configuração”, Mendes (2007). Segundo Mendes a ferramenta BigQuery, em especial, se destacou pelo suporte em relação aos dados transacionais e por possibilitar que a equipe não tivesse mais preocupação com o *setup* de infraestrutura de máquinas, períodos de manutenção e com disponibilidade imediata.

A tecnologia e a conectividade estão cada vez mais nos dando acesso a grandes quantidades de dados, neste caso, o maior desafio é saber filtrar e focar nas informações que realmente importam, afinal os dados são inúteis se não puderem ser transformados em inteligência. *Dashboard* é um termo que pode ser traduzido como "Painel de Controle". Os painéis têm todas as informações que é preciso em um só lugar, assim, é possível ter uma visão global do negócio para tomar as decisões certas. A parte boa é que o *dashboard* é uma ferramenta muito flexível, com ele, você pode selecionar dados relacionados e acompanhá-los. Além disso, os *dashboards* costumam gerar relatórios automatizados que trazem gráficos e imagens relacionados aos objetivos.

Um site que utiliza o processo de *dashboard* e o site de São Paulo contra a corona vírus, tendo como objetivo trazer informações que ajudem as pessoas a ter noção de

como anda as informações referente a pandemia, essas informações são do que diz respeito ao número de confirmados e óbitos, leitos e internações entre outras informações. Todas essas informações são mostradas em gráficos para deixar a visualização mais fácil para o usuário. Os dashboards tornam o processo de tomada de decisão cada vez mais fácil, centralizando informações de fácil visualização e compreensão, proporcionando uma visão ampla do assunto.

## 4. Desenvolvimento

### 4.1. Levantamento junto aos profissionais de saúde

No intuito de minimamente direcionar os dados a serem trabalhados pela ferramenta proposta, foi realizado um questionário simples junto a profissionais de saúde com a finalidade de gerar dados quantitativos e qualitativos a fim de detectar potenciais deficiências nos atuais canais de divulgação de informação a respeito da pandemia. Considerando o tempo e acesso disponível a estes profissionais, foi possível obter a resposta de 3 profissionais da saúde.

Foi levantado junto a esses profissionais a seguintes informações:

1. Função exercida no hospital?
2. O que poderia ajudar com as dúvidas frequentes dos pacientes referentes a pandemia?

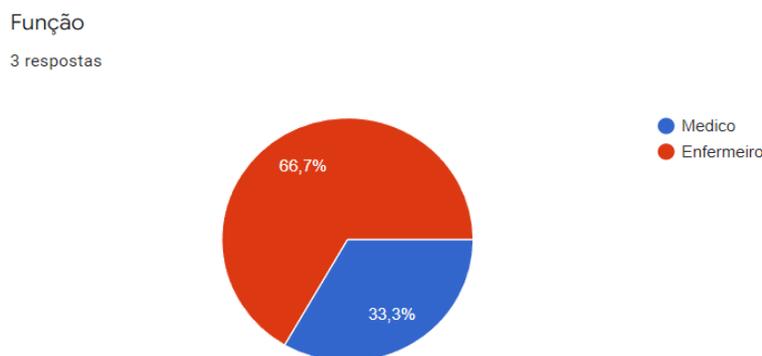


Figura 2. Esse gráfico nos mostra qual a função exercida pela pessoa da pesquisa

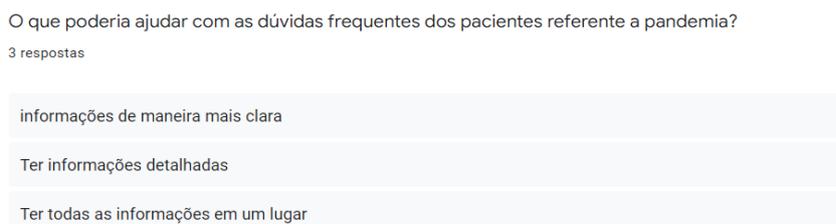


Figura 3. Nessa figura nos mostra a opinião das pessoas que responderam ao questionário.

Mesmo considerando a limitação de relevância estatística da pesquisa realizada, as respostas obtidas focam no mesmo escopo de problema. Em um momento posterior, para validar as respostas junto aos profissionais da área da saúde, todos apontaram que é comum que os pacientes apresentem dúvidas com relação a situação da pandemia, principalmente por ela não estar disponível de forma clara, detalhada e unificada.

#### 4.2. Análise das fontes de informações disponíveis

Com base nesse *feedback*, se fez uma análise do que seriam as principais fontes disponíveis para a população e como estavam sendo apresentadas.

Nesse sentido, foram identificados 3 sites: Coronavírus Brasil, OMS e SP contra o coronavírus.

Cada um dos sites foi analisado com relação aos seguintes quesitos: usabilidade, visualização e escopo dos dados disponíveis. O resultado desta análise pode ser observado na Tabela 1, onde foram listados os três sites que contém informações referente a pandemia com apresentações utilizando *dashboard*.

O primeiro a ser analisado foi o *dashboard* da corona vírus brasil, foi observado que a usabilidade é fácil e prática, a parte de apresentação dos dados é bem clara contudo os dados unificados não se enquadram, por só conter informações de estados e não cidades.

O segundo *dashboard* observado e analisado foi da OMS, a sua usabilidade não é intuitiva, ou seja, é bem confuso as informações, mais a visualização dos dados é apresentado, contudo não contém unificação dos dados por conta de só exibir informações dos países, não contendo informações de cidades nem de estados.

O terceiro e último *dashboard*, foi observado que não contém uma boa usabilidade tornando difícil a exploração pelo *dashboard*, ou seja é muito difícil navegar no *dashboard* e as informações são confusas, contudo apesar de ser confuso ainda apresenta os dados, já na parte de unificação dos dados, o *dashboard* exibe informações de alguns outros países que são Itália, Brasil e EUA, e de algumas cidades contudo, não contém o item vacinados. O terceiro *dashboard* se mostrou melhor no quesito informações que os demais *dashboards*.

Sites com dados da pandemia	Usabilidade	Visualização	Dados Unificados
Coronavírus Brasil	✓	✓	✗
OMS	✗	✓	✗
SP contra o coronavírus	✗	✓	✓

**Tabela 1. Comparativo dos Sites que contém informações referente a pandemia com apresentações de *dashboard*.**

Analisando a tabela, é possível observar que apenas um site apresenta uma boa usabilidade e dois deles apresentam um bom nível de visualização das informações,

nenhum deles atende a demanda inicial levantada junto aos profissionais de saúde em relação a informação da doença de forma consolidada.

### 4.3. Desenvolvimento do processo ETL e ambiente disponibilizado

No processo de ETL (Extração, transformação e carregamento dos dados) quantitativo, foram utilizadas as seguintes ferramentas:

- *BigQuery*, escolhida devido ao seu sistema interativo de indexação que permite consultas rápidas e completas. Ele é baseado na nuvem e possibilita diversas integrações com ferramentas ETL, assim facilitando nas buscas e coletas de dados referente a pandemia
- *Google Cloud*, após a coleta dos dados ter sido realizada, devido ao nível de integração com o *BigQuery*, esta ferramenta de armazenamento é uma solução natural para armazenar as informações tratadas e consolidadas das bases pesquisadas

O processo está sendo dividido em três módulos, tais demonstrados na seção 4.1. Desenho de Solução Implementada, Figura 2. Sendo eles dados brutos, dados refinados e visualização.

- **Dados Brutos:** Neste modulo são extraídos os dados de fontes oficiais em formato CSV, esses arquivos são armazenados manualmente no google cloud storage, na base de dados DadosCovid, apenas a fonte de dados de Hortolândia não foi incluída no cloud storage, por não conter dados públicos em CSV ou qualquer outra fonte. Foi preciso realizar uma coleta manualmente, inserindo diretamente nas tabelas do bigquery.
- **Dados Refinados:** após conter uma fonte única dos dados no cloud storage foi utilizado o *bigquery* para criação de tabelas refinadas, denominadas como CasosCovid, Estado, Municipios, VacinadosHorto, Internações e Óbitos. Para efeito de análise e aplicações de regras foram criadas *querys* para serem utilizadas na ferramenta *Data Studio* para a geração do *dashboard*.
- **Visualização:** com as *querys* prontas na etapa anterior foram criados planilhas e gráficos para apresentar de forma unificada e centralizadas dos dados extraídos de fontes diferentes. Este *dashboard* é responsivo para quaisquer dispositivos eletrônicos tais como, computadores, tablets e celulares. Além disso foi gerado uma URL para ser incorporado em qualquer site ou visualização de usuário.

#### 4.4. Proposta da Solução Implementada

A Figura 2 irá demonstrar a descrição das etapas do processo de forma resumida e clara, assim facilitando o entendimento de cada processo que foi realizado dentro de cada ferramenta que foi utilizada para se criar o projeto.

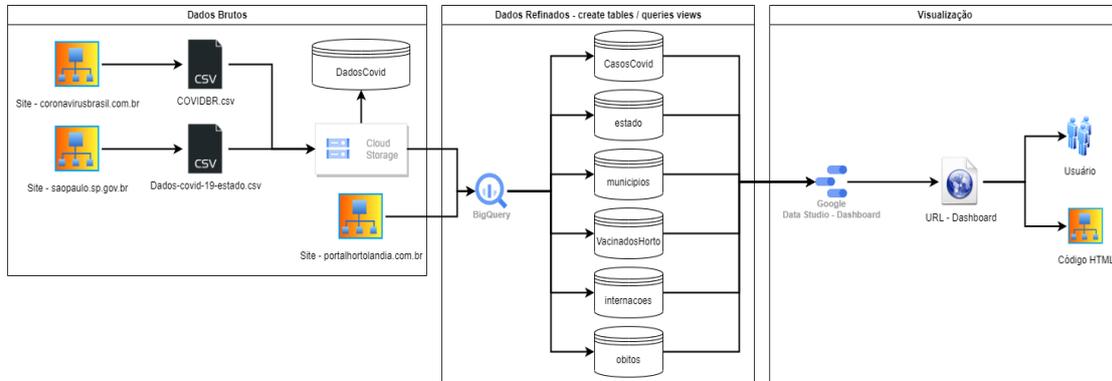


Figura 2. Desenho de Solução Implementada. Fonte: Elaborado pelo autor.

#### 4.5. MER (Modelo Entidade Relacionamento)

Segundo Martins (2007), em março de 1976 Peter P. Chen publicou um livro intitulado "The Entity-Relationship Model" no qual definiu uma possível abordagem ao processo de modelagem de dados. Passado a ser considerado a referência definitiva para dados ao longo dos anos, o processo de modelagem evoluiu para abordagens mais próximas de um ambiente orientado a objetos, que enriquece a técnica e, portanto, é adequado para novos propósitos.

O modelo do banco de dados utilizará o MER (Modelo Entidade Relacionamento). Modelagem e documentação do banco, organizando assim as tabelas e entendendo o detalhamento total do armazenamento de dados, como mostrado na Figura 3.

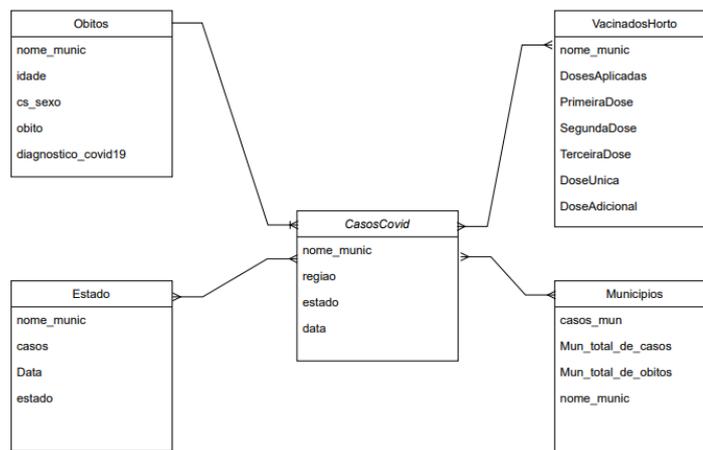


Figura 3. Modelo Entidade Relacionamento.

## 4.6. Caso De Uso

O diagrama de casos de usos irá descrever os requisitos funcionais do sistema de maneira consensual entre o usuário e o sistema. “É necessário frisar aqui, que um diagrama de caso de uso não tem reflexo direto em codificação. Ele é um instrumento de visualização de complexidade, é a abstração da complexidade”. (Medeiros, p.28). Os seus elementos principais são nome do caso de uso e os autores envolvidos no sistema. A Figura 4 , nos mostra o diagrama de casos de uso do processo proposto.

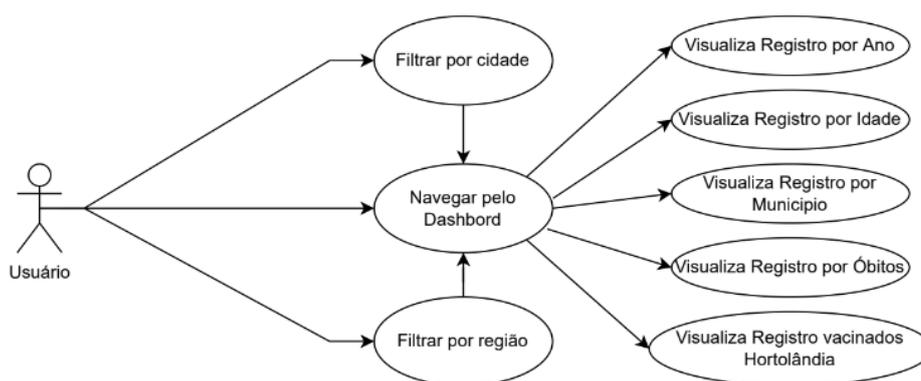


Figura 4. Diagrama de Casos de Uso.

## 4.7. Especificação de Requisitos

Para termos uma visão mais detalhada do *software* implementado, a melhor alternativa é as especificações dos requisitos como um todo.

“Fazer um bom levantamento e uma especificação de requisitos é algo primordial para quem trabalha com desenvolvimento de sistemas. Esse levantamento pode não garantir que o *software* contemple todas as reais necessidades dos usuários, mas tende a antecipar o surgimento dos erros de entendimento e inconsistências, aprimorando o processo de desenvolvimento de produtos de *softwares*.” (Machado, 2016, p.06).

### 4.7.1. Requisitos Funcionais

Os requisitos funcionais são requisitos de sistema de software que especifica uma função que o sistema ou componente deve ser capaz de realizar, ou seja, os requisitos que definem o comportamento do sistema, eles capturam as funcionalidades sob o ponto de vista do usuário.

- Gráfico por idade
- Gráfico por região
- Gráfico por mortes
- Gráfico por infectados
- Gráfico por ano

#### 4.7.2. Requisitos Não Funcionais

Requisitos não funcionais são aqueles que não estão diretamente relacionados a funcionalidades de um sistema. Segundo Sommerville (2007, p.82), os requisitos não funcionais podem estar relacionados às propriedades emergentes dos sistemas, como funcionalidade, tempo de resposta e espaço de armazenamento.

- O *dashboard* deve ser de fácil acesso
- As informações devem ser transmitidas de forma simplificada

### 5. Tratamento dos dados

#### 5.1. Complicações Encontradas

Na fase de buscas dos dados houve algumas complicações, nem todos os dados encontrados se encontravam em formato CSV, e os dados encontrados não estavam unificados como consequência houve um maior tempo gasto para encontrar os dados necessários para serem trabalhados e tratados.

Outra complicação encontrada foi no reconhecimento do mapa de bolha onde não estava reconhecendo as regiões determinadas ocasionando um tempo considerável para poder detectar onde estava o erro.

Nas consultas, ou seja, nas *queries* desenvolvidas houve erros de sintaxe com isso, acabou atrasando um pouco, contudo consegui tratá-los a tempo, acredito que como acabou sendo uma ferramenta nova para mim trabalhar demorou um tempo para me adaptar, mas no final consegui me adaptar, sendo assim alguns dos erros que houve foram também por conta disso, contudo todos os erros foram tratados.

### 6. Resultados

O processo desenvolvido, ETL(Extração, transformação e carregamento dos dados), tem como objetivo ajudar as pessoas a terem uma melhor visão da Covid ou seja, como devemos nós cuidar, quantidades de vacinados, quantas de pessoas que já pegaram a doença, entre outras... Essas informações podem ajudar com a conscientização das pessoas ao verem os dados apresentados em referência a pandemia.

#### 6.1. Tela Detalhamento *Dashboard*

Como mostrado na Figura 5, quando inicializado o *dashboard* o usuário é direcionado

para a tela de detalhamento do *dashboard*, onde mostra detalhes de cada tela direcionando usuário com links para as demais telas.

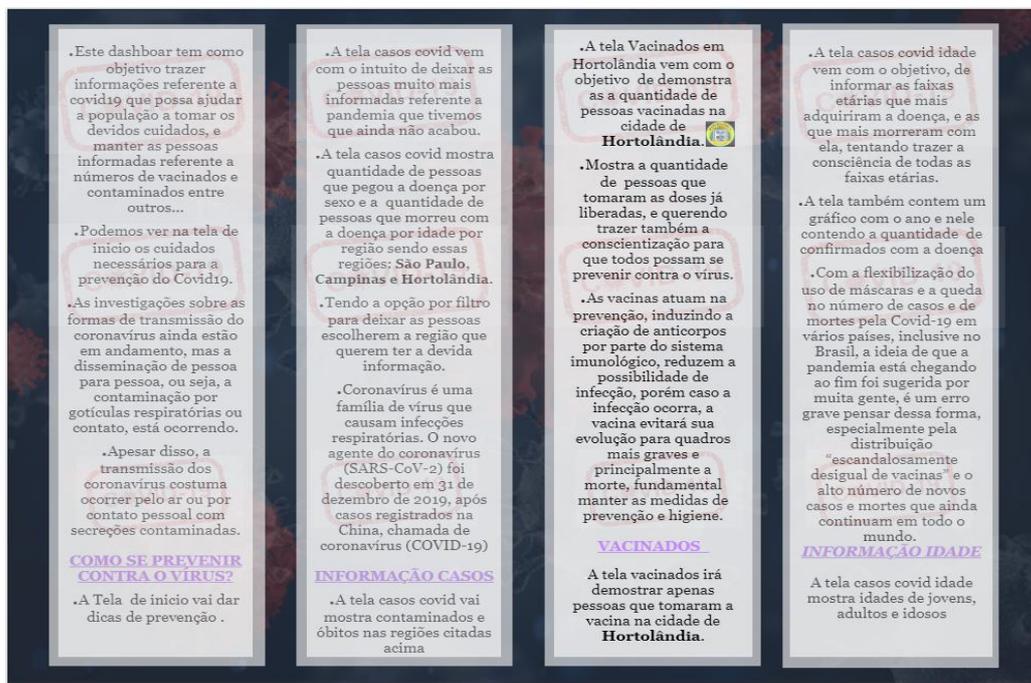


Figura 5. Tela Inicial.

## 6.2. Tela Inicial do Dashboard

Como mostrado na Figura 6, a tela de início do *dashboard* mostra detalhes de como se prevenir contra a covid fazendo os usuários evitar a proliferação do vírus.

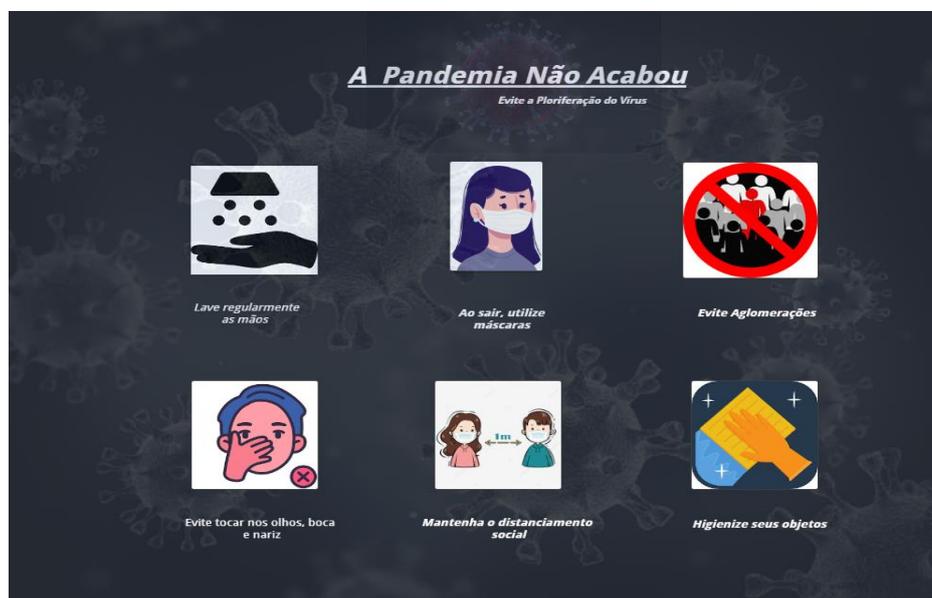


Figura 6. Tela Inicial.

### 6.3. Tela Casos Covid

Como mostrado na Figura 7, o usuário pode ter acesso aos números de casos de covid nas regiões de São Paulo, Campinas e Hortolândia.



Figura 7. Tela Casos Covid.

### 6.4. Tela Casos Covid Idade

Como mostrado na Figura 8, o usuário pode ter acesso aos números de casos de covid por idade nas regiões de São Paulo, Campinas e Hortolândia, mostrando a faixa etária que mais se contaminou.

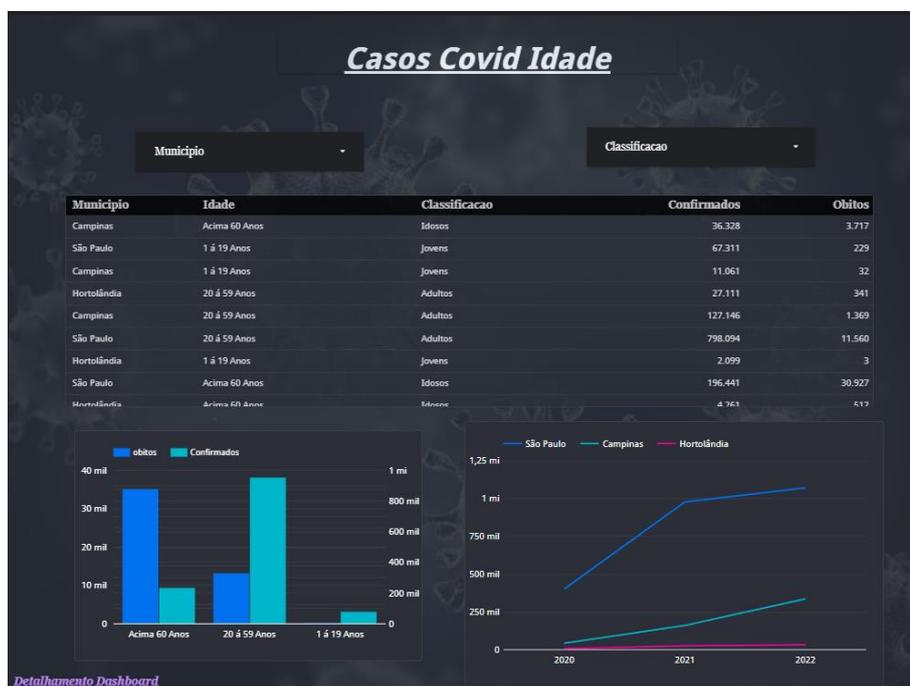


Figura 8. Tela Casos Covid Idade.

## 6.5. Tela Vacinas Hortolândia

Como demonstrado na Figura 9, o usuário pode ter acesso a quantidade de pessoas imunizadas que tomaram as doses corretamente.

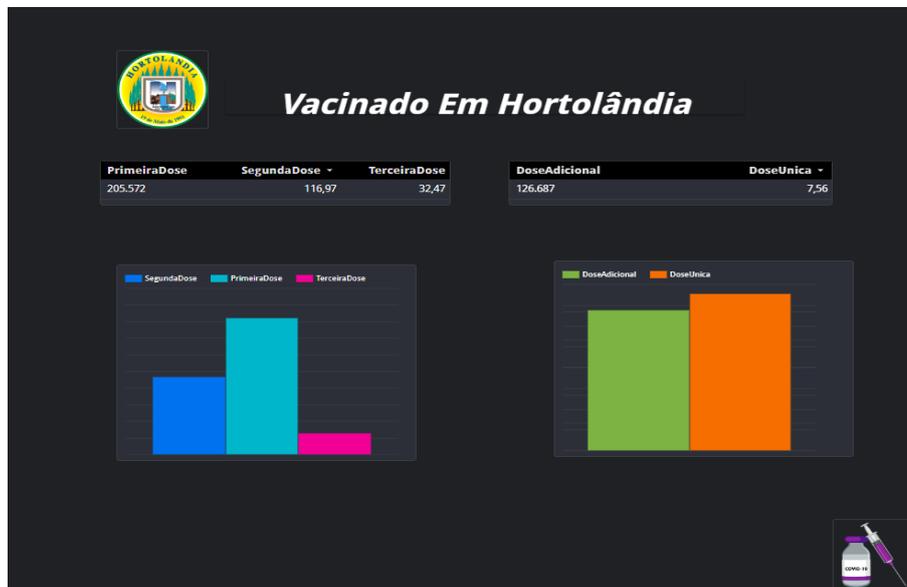


Figura 9. Tela Vacinados em Hortolândia.

## 7. Teste de Usabilidade

Para obter uma melhor avaliação sobre um produto ou serviço, em termos de facilidade e intuição, para melhores resultados desta técnica, foram realizados com usuários que representam personas, para quem o produto é destinado, ou seja, o teste foi realizado com três trabalhadores da saúde sendo eles dois enfermeiros e um médico, que deram notas de 1 a 10, para que pudéssemos observar algum eventual problema ou possibilidade de melhorias.

<i><b>Pesquisa</b></i>	
<i>Enfermeiro 1</i>	<ul style="list-style-type: none"> <li>● Usabilidade:9,0</li> <li>● Conteúdo com retorno pertinente :9,0</li> <li>● Auxílio nas informações:9,5</li> </ul> <p><b>Melhorias:</b> Mostrar quantidade de pessoas que pegaram novamente a doença</p>
<i>Enfermeiro 2</i>	<ul style="list-style-type: none"> <li>● Usabilidade:9,5</li> <li>● Conteúdo com retorno pertinente :8,5</li> <li>● Auxílio nas informações:9,0</li> </ul> <p><b>Melhorias:</b> Mostrar quantidade de pessoas hospitalizadas</p>
<i>Médico</i>	<ul style="list-style-type: none"> <li>● Usabilidade:9,0</li> <li>● Conteúdo com retorno pertinente :8,5</li> <li>● Auxílio nas informações:8,5</li> </ul> <p><b>Melhorias:</b> Ter informações de outras cidades</p>

Tabela 2. Resultado da pesquisa

## 8. CONCLUSÃO

Com base no que foi apresentado neste trabalho, foi alcançado o objetivo de aplicar e desenvolver os conteúdos vistos durante o curso na proposta e desenvolvimento de uma solução para um problema prático e atual. No caso o processo de desenvolvimento ETL (Extração, transformação e carregamento dos dados) que resultou no desenvolvimento de um *dashboard* para visualização e customização de informações consolidadas a respeito dos casos de covid na região alvo.

Reforçamos que apesar dos resultados e análises serem direcionadas para validar a proposta apresentada, ela já demonstra que se desenvolvida ela contribuiria com a melhora no acesso da população a informações a respeito da Covid, tanto no aspecto de visualização de todo o contexto das informações disponíveis, quanto na facilidade de acesso as mesmas.

Considerando a conclusão de todas as etapas inicialmente planejadas, ao final do processo foi proposto ainda um teste de usabilidade, procurando inserir no processo o *feedback* dos usuários que participaram do processo, obtendo informações a respeito do nível de satisfação observado e dados a respeito de melhorias a serem incorporadas em novas iterações do desenvolvimento da ferramenta proposta.

De forma mais específica, o referente trabalho de conclusão de curso, demandou conhecimento de diversas áreas. A utilização de conhecimento na linguagem SQL foi responsável para o refinamento dos dados realizado dentro do *BigQuery*. As habilidades desenvolvidas em disciplinas como lógica de programação e Banco de dados (BD), foram importantes no processo de seleção dos dados nas diversas bases estudadas. O conteúdo visto em engenharia de software contribuiu com o planejamento e proposta das funcionalidades da aplicação e planejamento das fases de desenvolvimento do projeto e o uso da metodologia ETL. Um dos desafios foi a o processo de lidar com bases de origens diferentes e a utilização de algumas tecnologias como o *BigQuery* que não foram abordadas diretamente durante o curso. Contudo, como sua base está nos conceitos de BD e utiliza a linguagem era em SQL, e outras habilidades desenvolvidas nos demais componentes do curso, a composição de todos esses conhecimentos contribuiu na finalização com êxito do projeto.

## 9. TRABALHOS FUTUROS

Em termos de trabalho futuro, existem dois caminhos a serem seguidos:

1. O primeiro caminho a ser desenvolvido, será a migração dos dados em CSV para uma API (Application Programming Interface), as APIs, consistem em um conjunto de rotinas e padrões de programação para acesso a um aplicativo de software ou plataforma baseado na Web, a mesma pode se ligar em diversas funções em um site de maneira que possam ser utilizadas em outras aplicações. Com base nesta funcionalidade, o processo de inserir dados será mais fácil pois os dados serão atualizados direto da origem sem intervenção humana. Além de algumas vantagens como, redução a manutenção do código, e facilidade na hora de realizar a atualização dos dados.

2. O segundo caminho seria a inserção de novas cidades pelo tempo ser bem escasso só foi inserido informações de três cidades, contudo nada impede de no futuro ser inserido mais cidades, quanto mais informação melhor ficara o *dashboard*, os dados

já se encontram unificados facilitando assim o trabalho de se inserir novas informações.

## REFERÊNCIAS

Big Data: O que é, conceito e definição. **Cetax** (2022). Disponível em:  
<https://www.cetax.com.br/blog/big-data/>

*BigQuery*. **Google Cloud** (2021).

Disponível em:

[https://cloud.google.com/BigQuery?utm\\_source=google&utm\\_medium=cpc&utm\\_campaign=latam-BR-all-pt-dr-SKWS-all-all-trial-p-dr-1011454-LUAC0015653&utm\\_content=text-ad-none-any-DEV\\_c-CRE\\_532282943216-ADGP\\_Hybrid%20%7C%20SKWS%20-%20PHR%20%7C%20Txt%20~%20Data-Analytics\\_BigQuery-KWID\\_43700064907853690-kwd-332753161828&utm\\_term=KW\\_BigQuery-ST\\_BigQuery&gclid=Cj0KCQiAoY-PBhCNARIsABcz771BBgw8wPO-1rFJRabUBOk917vZXuSptVNRAU-V31nR2lZ-dQXgyIkaAuEvEALw\\_wcB&gclsrc=aw.ds](https://cloud.google.com/BigQuery?utm_source=google&utm_medium=cpc&utm_campaign=latam-BR-all-pt-dr-SKWS-all-all-trial-p-dr-1011454-LUAC0015653&utm_content=text-ad-none-any-DEV_c-CRE_532282943216-ADGP_Hybrid%20%7C%20SKWS%20-%20PHR%20%7C%20Txt%20~%20Data-Analytics_BigQuery-KWID_43700064907853690-kwd-332753161828&utm_term=KW_BigQuery-ST_BigQuery&gclid=Cj0KCQiAoY-PBhCNARIsABcz771BBgw8wPO-1rFJRabUBOk917vZXuSptVNRAU-V31nR2lZ-dQXgyIkaAuEvEALw_wcB&gclsrc=aw.ds)

Coronavírus Brasil. **Covid.Saude** (2022). Disponível Em:  
<https://covid.saude.gov.br/>

Criando Diagramas UML com o StarUML. **CNX.ORG** (2013). Disponível em:  
[https://cnx.org/contents/sKehW\\_Tl@1/Criando-Diagramas-UML-com-o-StarUML](https://cnx.org/contents/sKehW_Tl@1/Criando-Diagramas-UML-com-o-StarUML)

Cloud Computing Significado. **Mandic** (2022). Disponível Em:  
<https://www.mandic.com.br/cloud/>

Data Studio. **Datastudio** (2022). Disponível em:  
<https://datastudio.google.com/u/0/>

Dotz. **Google Cloud** (2007). Disponível Em:  
<https://cloud.google.com/customers/dotz/?hl=pt-br>

Engenharia de Dados: Tudo que você precisa saber. **FIA** (2019). Disponível Em:  
<https://fia.com.br/blog/engenharia-de-dados/#:~:text=nas%20m%C3%A3os%20corretas-,O%20que%20%C3%A9%20Engenharia%20de%20Dados%3F,grande%20quantidade%20de%20informa%C3%A7%C3%A3o%20dispon%C3%ADvel.>

Google Forms o que é e como usar. **Techtudo** (2018). Disponível em:  
<https://www.techtudo.com.br/dicas-e-tutoriais/2018/07/google-forms-o-que-e-e-como-usar-o-app-de-formularios-online.ghtml>

Google Data o que é e Como Usá-lo. **Rockcontent** (2022). Disponível em:  
<https://rockcontent.com/br/blog/google-data-studio/>

Hortolândia inicia vacinação. **Hortolandia** (2022). Disponível em:  
<http://www2.hortolandia.sp.gov.br/noticias/itemlist/tag/4%C2%AA%20dose>

Mer/Der. **Alura** (2022). Disponível em:  
<https://www.alura.com.br/artigos/mer-e-der-funcoes>

O Antes Durante e Depois de um Processo ETL. **Etz** (2020). Disponível em:  
<https://etz.com.br/entenda-o-processo-de-etl/>

O que é o BigQuery e quais são os seus principais benefícios? **DigitalHorse** (2022).  
Disponível Em: <https://www.digitalhouse.com/br/blog/bigquery/>

O que é ETL?. **Oracle** (2022). Disponível em:  
<https://www.oracle.com/br/integration/what-is-etl/>

O que é SQL e para que ele Serve?. **TecMundo** (2019). Disponível em:  
<https://www.tecmundo.com.br/software/146482-sql-que-ele-serve.htm>

O que é e como usar os principais comandos básicos SQL. **Betrybe** (2022).  
Disponível Em: <https://blog.betrybe.com/sql/>

Painel da OMS sobre o coronavírus (COVID-19). **World Health Organization** (2022).  
Disponível Em: <https://covid19.who.int/>

Para que Serve o Google Cloud Platform e Quais Vantagens. **Multiedro** (2021).  
Disponível em: <https://blog.multiedro.com.br/para-que-serve-o-google-cloud->

SP Contra o Corona Vírus. **Governo do Estado de São Paulo** (2022). Disponível Em:  
<https://www.seade.gov.br/coronavirus/>

# Documento Digitalizado Público

## Artigo TCC - Elizabeth dos Santos

**Assunto:** Artigo TCC - Elizabeth dos Santos  
**Assinado por:** Edgar Noda  
**Tipo do Documento:** Dissertação  
**Situação:** Finalizado  
**Nível de Acesso:** Público  
**Tipo do Conferência:** Documento Digital

Documento assinado eletronicamente por:

- **Edgar Noda, PROFESSOR ENS BASICO TECN TECNOLOGICO**, em 14/11/2022 17:54:40.

Este documento foi armazenado no SUAP em 14/11/2022. Para comprovar sua integridade, faça a leitura do QRCode ao lado ou acesse <https://suap.ifsp.edu.br/verificar-documento-externo/> e forneça os dados abaixo:

**Código Verificador:** 1157507

**Código de Autenticação:** b3489c5929

