

## DESVENDANDO O ENADE: UMA ANÁLISE ABRANGENTE DOS DADOS E RESULTADOS DOS CURSOS SUPERIORES.

### DESVENDANDO O ENADE: UMA ANÁLISE ABRANGENTE DOS DADOS E RESULTADOS DOS CURSOS SUPERIORES.

920

Juliano Alessandro dos Santos<sup>1</sup>; Mateus Guilherme Fuini<sup>2</sup>

1 - Discente em Desenvolvimento de Software Multiplataforma pela Faculdade de Tecnologia de Itapira “Ogari de Castro Pacheco” (FATEC de Itapira); 2 - Mestre em Engenharia da Computação pela Universidade Estadual de Campinas. Docente da FATEC de Itapira.

**Contato:** mateus.fuini@fatec.sp.gov.br

#### RESUMO

Este estudo, referente à área de engenharia de dados, tem como objetivo realizar a extração, processamento e análise de dados dos cursos superiores, permitindo dessa maneira auxiliar na compreensão de tendências e insights. Em virtude disso, é válido notar a relevância desses estudos, pois com ele é possível vislumbrar uma análise mais quantitativa dos dados, ou seja, quais foram as notas mais altas por instituições e idade. Para poder atingir tal objetivo, todo o estudo será contemplado com uma metodologia descritiva dos dados, ou seja, iremos recolher as informações e apresentar todos esses dados. Em virtude disso, serão apresentados os resultados de como foram feitos os processos de extração e limpeza, além de apresentar um painel de dados no qual, demonstra todos os dados cruzados organizados e em formato de gráfico, com filtro de tempo para compreender ainda melhor essa dinâmica. Portanto, podemos inferir que os resultados foram obtidos e que os limites dos meus estudos são pautados pelas limitações dos dados, no qual não foi possível cruzar as informações de cota social, e bolsista.

**Palavras-chave:** exportação; processamento; limpeza; análise; padrões

#### ABSTRACT

This study, related to the field of data engineering, aims to extract, process, and analyze data from higher education courses, thus aiding in the understanding of trends and insights. It is important to note the relevance of these studies, as they allow for a more quantitative analysis of the data, such as identifying which institutions have the highest scores and correlating this information with student age. To achieve this goal, the study will follow a descriptive data methodology, collecting and presenting all relevant information. As part of this, the study will detail the processes of data extraction and cleaning, and it will also feature a data dashboard that organizes and cross-references all collected information in graphical form, with time filters to better understand the dynamics at play. Therefore, we can infer that the results have been obtained, though the study is limited by the constraints of the available data, which did not allow for cross-referencing information on social quotas and scholarship recipients.

**Keywords:** extraction; processing; cleaning; analysis; patterns

## 1. INTRODUÇÃO

O acesso ao ensino superior, é uma grande porta de entrada para a mudança de vida e melhora na qualidade de vida dos Brasileiros, em conformidade com ministério da Educação(2022), o acesso à educação superior em universidades públicas alcançaram valores de 24.6% em 2018. Além disso, pesquisas realizadas pelo Instituto Brasileiro de Geografia e Estatística(IBGE), em conformidade com Bello e Brito(2024), é apresentado que em 2022, os estudantes que estavam no ensino superior passaram a ser de 21.6%, demonstrando uma queda de entrada de estudantes no ensino superior de 3%, se comparado com 2018.

Diante de tal fato, emergiu-se a necessidade do governo de ranquear as melhores universidades do Brasil, para que os estudantes possam escolher as universidades que oferecem maior qualidade educacional. Em virtude disso, é proposto a metodologia chamada ENADE. Segundo a Empresa Parceira (2022), o Exame Nacional de Desempenho dos Estudantes (ENADE) tem como propósito medir a qualidade dos cursos superiores, além de ser uma ferramenta valiosa para os futuros ingressantes, o ENADE também beneficia as instituições de ensino. Com ele, é possível determinar o nível educacional dos alunos, proporcionando uma avaliação detalhada da instituição. Isso pode destacar as universidades que se sobressaem em termos de qualidade educacional.

Nessa linha de pensamento, é válido pensar quais instituições conseguem obter as maiores notas desse exame, qual a faixa etária predominante dessas notas altas e qual o sexo das notas mais altas registradas, tais questionamentos serão abordados mais para frente na sessão referente aos resultados e discussões.

Para nortear toda a construção do referido projeto foram levantadas algumas perguntas que possibilitam centralizar a referente monografia. Desse modo, em conformidade com o parágrafo anterior pode-se deduzir algumas perguntas norteadoras sendo elas:

- Ao final do projeto é possível identificar padrões e tendências nos dados cruzados?
- A construção de um painel de visualização foi alcançada com sucesso?

- Foi possível cruzar todos os dados tratados?

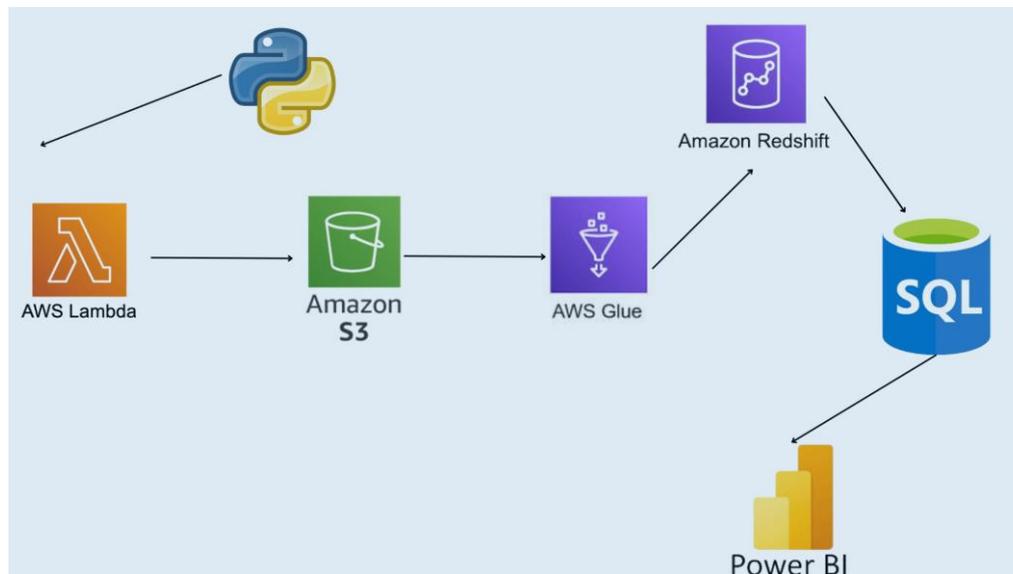
## 2. PROCEDIMENTOS METODOLÓGICOS

A metodologia que contempla essa iniciação científica é a quantitativa, visto que será analisado um conjunto de dados do site: Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira | Inep, com esses dados coletados e tratados iremos analisar os principais fenômenos relacionados a esses dados, e tentar compreender se tais comportamentos se mantêm e porquê tais procedimentos acontecem.

Os dados que serão analisados, foram obtidos através de um formulário que é passado para os estudantes responderem todos os anos. Nesse sentido, os dados ficam públicos no site citado anteriormente para qualquer indivíduo acessar e visualizar as informações.

Para coletar os dados, implementamos uma rotina de automação em Python para extrair os arquivos e armazenar esses dados em um bucket na AWS. Em seguida, desenvolvemos outra rotina para descompactar os arquivos zipados e transferi-los para outro bucket. Os dados foram então mapeados pelo crawler e posteriormente processados e limpados preliminarmente pelo AWS Glue e carregados no banco de dados Redshift. Por fim, foram transferidos para o SQL Server antes de serem migrados para o Power BI, onde foram analisados para insights valiosos. Na figura 3, é apresentado um resumo de todas as etapas que foram realizadas durante o projeto.

Figura 3 - Resumo do Projeto



Fonte: Próprio Autor (2024)

923

## 2.1 DEFINIÇÃO NUVEM

Em conformidade com Pereira et al(2011, apud BUYYA, 2008), sistemas em nuvem, é um tipo de serviço que consiste em ter um sistema distribuído e paralelo de computadores, que estão conectados entre si permitindo dessa forma a utilização destes recursos através de acordos de utilização de seus usuários. Nesse contexto, tais recursos que são consumidos pelos usuários são alocados em hardware que são alocados em data centers (PEREIRA et al.,SOUSA, 2009 ).

### 2.1.1 VANTAGENS EM UTILIZAR SERVIÇOS EM NUVEM

A seguir serão apresentadas algumas vantagens em se utilizar serviços em nuvem.

#### 2.1.1.1 SERVIÇOS SOB DEMANDA

O usuário que utiliza um serviço em nuvem, é capaz de gerenciar a quantidade de recursos de memória e tempo de processamento que o mesmo utiliza, ou seja, o mesmo vai adicionando tais recursos ao seu sistema conforme a escalabilidade de seu projeto cresça (PEREIRA et al., 2011).

### 2.1.1.2 INDEPENDÊNCIA DE LOCALIZAÇÃO

Como seus recursos estão centralizados em grandes data centers, sua utilização se dá pela conexão da internet com esses serviços, portanto ao se conectar você pode acessar seus recursos em qualquer localidade (PEREIRA et al., 2011).

### 2.1.1.3 MEDIÇÃO DE SERVIÇO

Em concordância com Pereira et al. (2011), os serviços em nuvem são contabilizados, na medida em que o usuário estiver utilizando tais recursos, no momento em que ele não utilizar o serviço não será cobrado.

### 2.1.2 AWS

A AWS (Amazon Web Services), é um serviço em nuvem desenvolvido pela Amazon, contendo em seu catálogo mais de 200 serviços que podem ser ofertados aos seus usuários. Nesse contexto, a plataforma adquire seus ganhos através do tempo e quantidade de serviços que o usuário utiliza (TECNOBLOG, 2024).

## 3. RESULTADOS E DISCUSSÕES

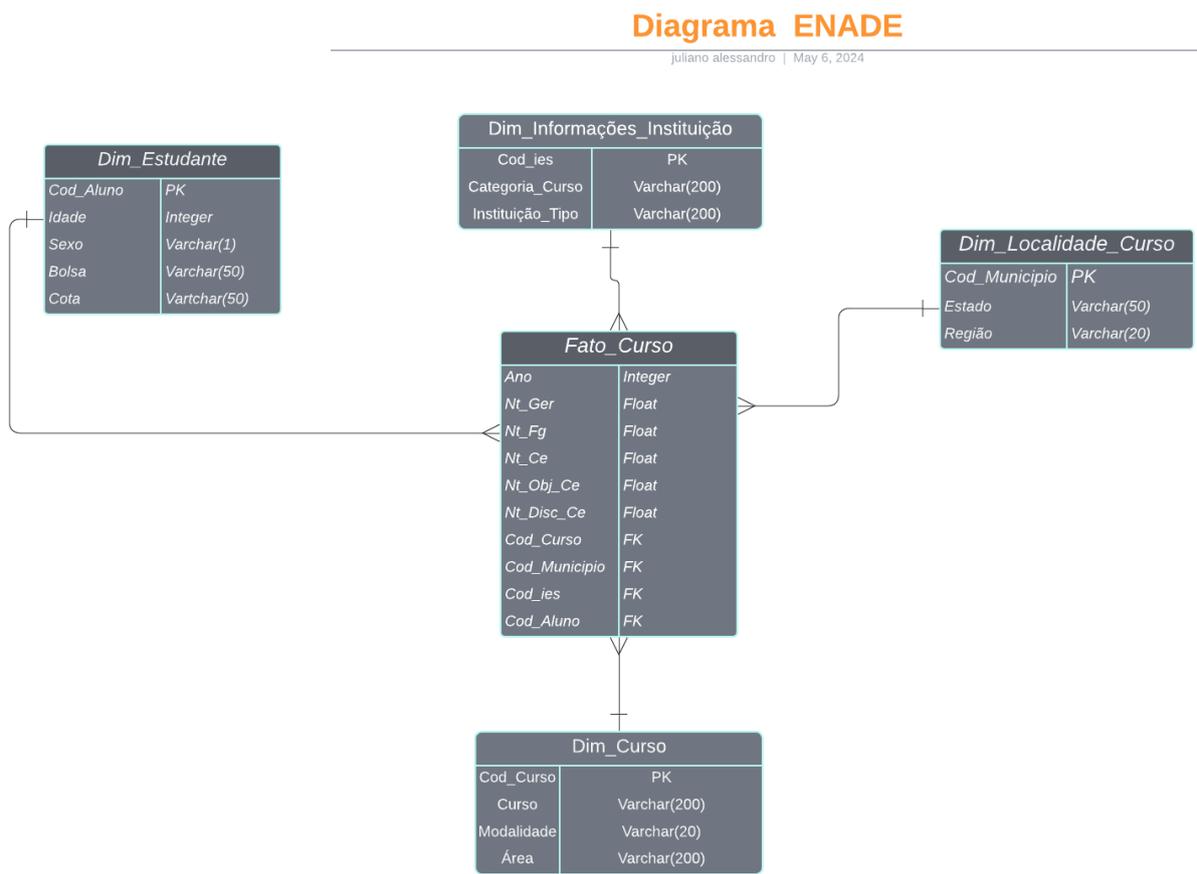
Neste capítulo serão abordados os principais resultados obtidos durante o desenvolvimento do projeto, além de mostrar algumas limitações e dificuldades que foram encontradas durante a realização do referido trabalho.

### 3.1 MODELAGEM ESTRELA

Após os dados serem tratados e processados pelas ferramentas da AWS, que serão citadas posteriormente, o resultado final será parecido com a figura 4. Nesse contexto, a figura abaixo está seguindo as diretrizes do modelo estrela, no qual consiste em um modelo multidimensional, em que as tabelas recebem duas identificações: tabela Dimensão, quando se refere a categorizar um assunto específico, ou seja, é a identificação dos elementos presentes naquela tabela, enquanto a tabela fato, se refere a coleção de eventos que ocorrem ao longo do tempo, no qual as tabelas dimensões participam (MACHADO, 2016, p.79 apud SOUZA et al., 2021, p. 36). Na figura 4, a identificação da tabela do tipo Dimensão se

dá pelo uso do prefixo DIM, enquanto a tabela fato é identificada pelo nome FATO. Nessa linha de pensamento é possível perceber que tal modelo segue um padrão no qual os relacionamentos só acontecem com a tabela fato, em virtude disso ganha-se no poder de processamento pois joins são pesados, então para otimizar consultas utiliza-se o modelo multidimensional, que prioriza velocidade durante as consultas.

Figura 4- Diagrama ENADE modelo Estrela

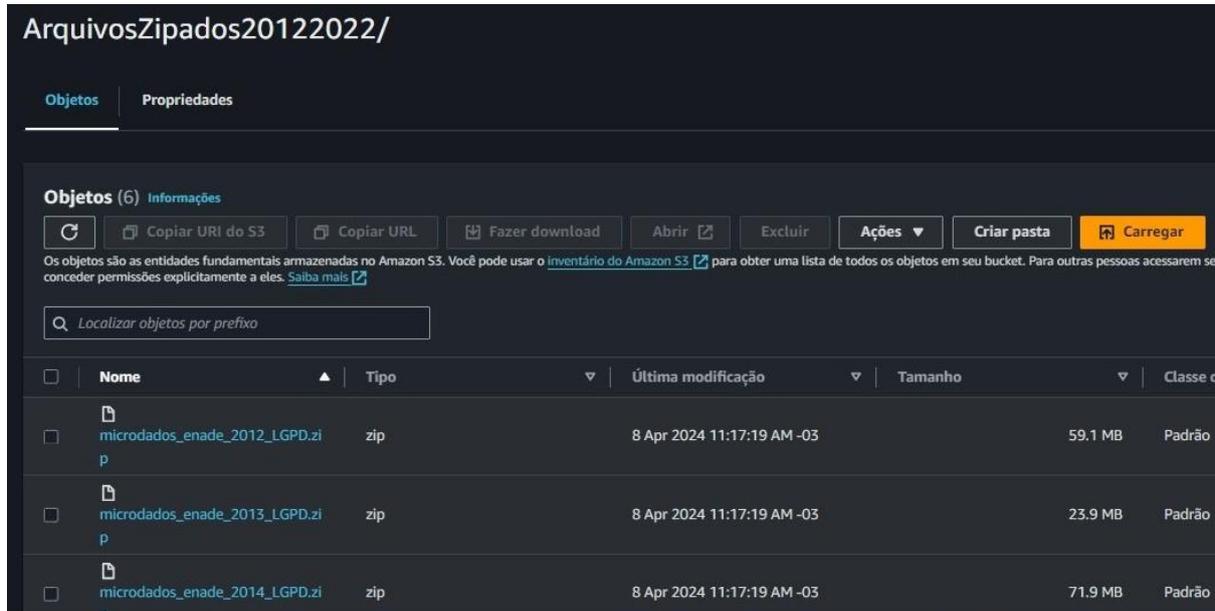


Fonte: Próprio Autor (2024)

### 3.2 BUCKET

Na primeira etapa do projeto foi necessário armazenar os 10 anos de arquivos zipados no bucket para posteriormente ser extraído seu conteúdo, assim como demonstrado na figura 5.

Figura 5 - Demonstração da utilização do bucket



Fonte: Próprio Autor (2024)

### 3.3 AUTOMAÇÃO DE EXTRAÇÃO DE ARQUIVOS ZIPADOS

Para poder extrair o conteúdo dos arquivos zipados e posteriormente encaminhar seu conteúdo para outro bucket específico, foi necessário criar uma função em Python que automatize todo esse processo citado anteriormente. Nesse contexto, a AWS, oferece um serviço chamado lambda, no qual permite ao usuário utilizar uma linguagem de programação de sua preferência para realizar determinadas rotinas. Nesse sentido, a figura 6 aborda toda codificação feita para poder extrair o conteúdo dos arquivos zipados e posteriormente encaminhar seu conteúdo para o outro Bucket, assim como é mostrado na figura 7. Dessa forma, podemos perceber pela figura 6, que foi utilizado algumas bibliotecas essenciais para realizar todo o processo citado anteriormente, a biblioteca json, foi utilizada para dar uma resposta se a nossa função está sendo respondida corretamente pelo servidor, a boto3, é uma biblioteca própria criada pela aws para poder se comunicar com os diversos serviços oferecidos pela plataforma, com isso sua utilização nesse contexto se deu para poder se comunicar com o bucket `arquivosextraidos20122021`, demonstrado na figura 7, com a biblioteca Bytes IO, podemos ler os bytes e converte em um arquivo temporário,

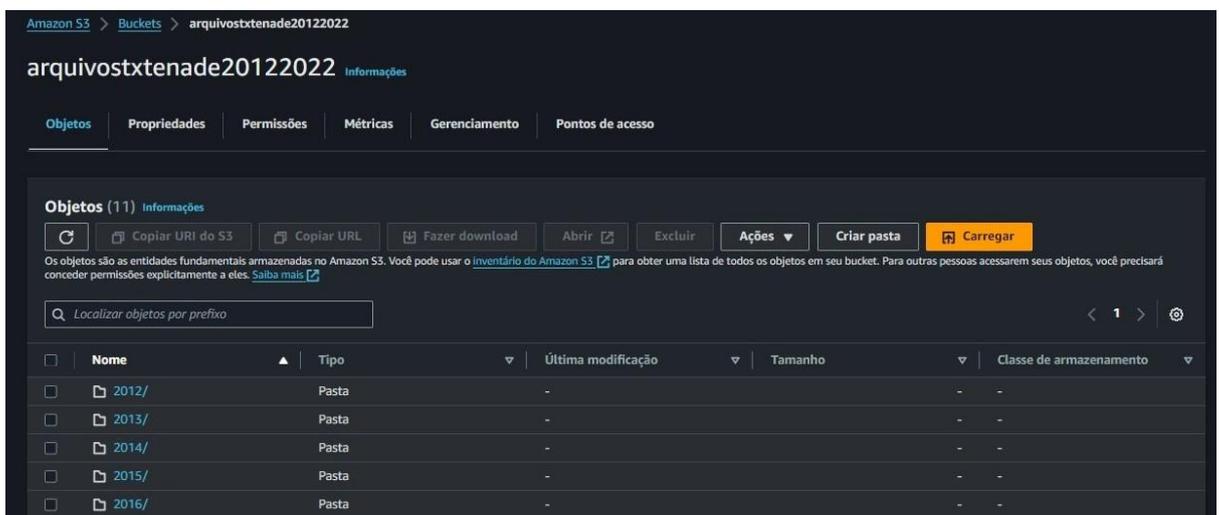
para que dessa forma com a biblioteca Zipfile, seja possível acessar o conteúdo dos arquivos zipados e posteriormente encaminhar todos os arquivos txt para seu lugar de destino.

Figura 6 - Demonstração do Código em Python para automação de rotinas

```
1 import json
2 import boto3
3 from io import BytesIO
4 import zipfile
5
6 def lambda_handler(event, context):
7     session = boto3.session.Session()
8     dev_cliente = session.client('s3')
9     dev_resource = boto3.resource('s3')
10    print("1")
11    bucket_dev = dev_resource.Bucket('dadoszipados')
12    zip_object = dev_resource.Object(bucket_name='dadoszipados', key='ArquivosZipados20122022/microdados_enade_2022_LGPD.zip')
13    print("2")
14    buffer = BytesIO(zip_object.get()["Body"].read())
15    z = zipfile.ZipFile(buffer)
16    print("3")
17    #codigo para aqui
18    for filename in z.namelist():
19        file_info = z.getinfo(filename)
20        print(f'Copiando arquivo {filename} para arquivostxtenade20122022')
21        response = dev_cliente.put_object(
22            Body=z.open(filename).read(),
23            Bucket='arquivostxtenade20122022',
24            Key=f'2022/{filename}'
25        )
26    return {
27        'statusCode': 200,
28        'body': json.dumps('Hello from Lambda!')}
29
```

Fonte: Próprio Autor (2024)

Figura 7 - Bucket de destino dos arquivos txt



Fonte: Próprio Autor (2024)

### 3.4 MAPEAMENTO DOS DADOS

Uma das etapas fundamentais que servirá de alicerce para a próxima etapa do projeto, é o mapeamento dos dados, utilizando-se o serviço Crawler da AWS é possível identificar quais são os tipos de dados presentes naquele arquivo, assim como saber quais colunas existem no meu arquivo, com essa etapa feita se torna mais fácil a manipulação de dados para processos de ETL. Com isso na figura 8, é apresentado os Crawlers criados para realizar o mapeamento dos dados, e na figura 9 é demonstrado quais tipos de dados e quais foram as colunas encontradas pelo Crawler.

Figura 8 - Crawler criados para o mapeamento dos dados

Name	State	Schedule	Last run	Last run time...	Log
<a href="#">CatalogarDados</a>	Ready		Succeeded	April 3, 2024 at 1...	<a href="#">View log</a>
<a href="#">CatalogoENADE2012Dados</a>	Ready		Succeeded	April 8, 2024 at 1...	<a href="#">View log</a>
<a href="#">CatalogoENADE2013Dados</a>	Ready		Succeeded	April 8, 2024 at 1...	<a href="#">View log</a>
<a href="#">CatalogoENADE2014Dados</a>	Ready		Succeeded	April 8, 2024 at 1...	<a href="#">View log</a>
<a href="#">CatalogoENADE2015Dados</a>	Ready		Succeeded	April 8, 2024 at 1...	<a href="#">View log</a>
<a href="#">CatalogoENADE2016Dados</a>	Ready		Succeeded	April 8, 2024 at 1...	<a href="#">View log</a>
<a href="#">CatalogoENADE2017Dados</a>	Ready		Succeeded	April 8, 2024 at 1...	<a href="#">View log</a>
<a href="#">CatalogoENADE2019Dados</a>	Ready		Succeeded	April 8, 2024 at 1...	<a href="#">View log</a>
<a href="#">CatalogoENADE2021Dados</a>	Ready		Succeeded	April 8, 2024 at 1...	<a href="#">View log</a>

Fonte: Próprio Autor (2024)

Figura 9 - Resultado do mapeamento de dados

#	Column name	Data type	Partition key	Comment
21	qe_i14	string	-	-
22	qe_i15	bigint	-	-
23	qe_i16	string	-	-
24	qe_i17	string	-	-
25	qe_i18	string	-	-
26	qe_i19	string	-	-
27	qe_i20	string	-	-
28	qe_i21	string	-	-
29	tp_sexo	string	-	-
30	nu_idade	bigint	-	-
31	qe_i01	string	-	-
32	qe_i02	string	-	-
33	qe_i03	string	-	-

Fonte: Próprio Autor (2024)

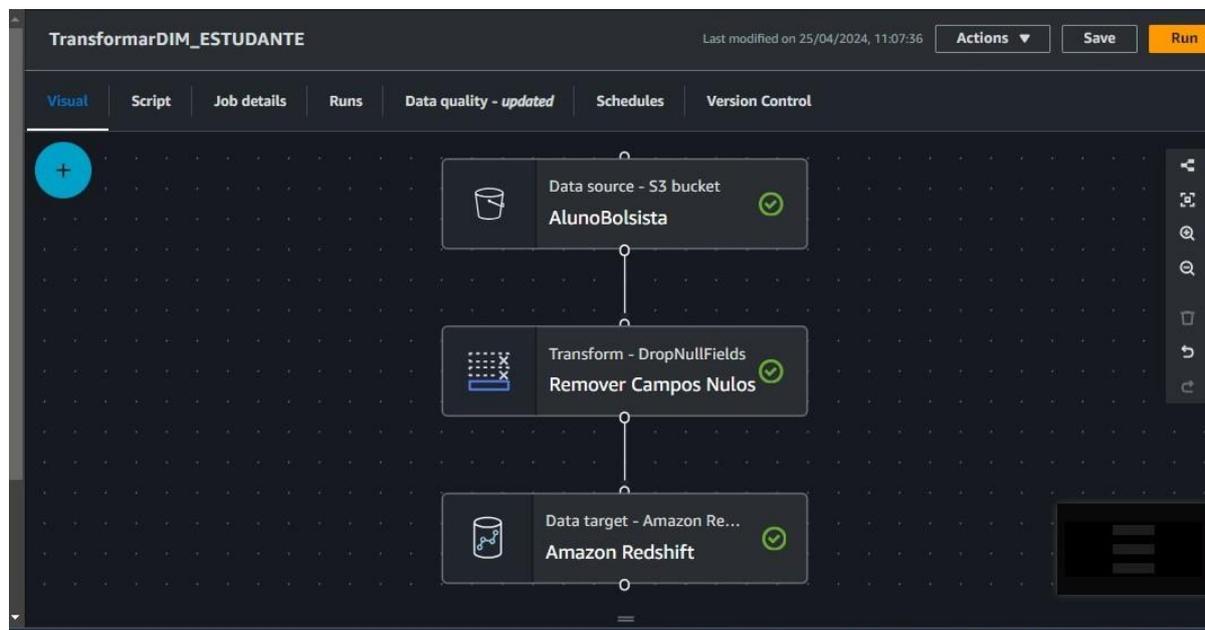
### 3.5 AWS GLUE

O Serviço do Glue da AWS, permite ao engenheiro de dados realizar uma limpeza nos dados. Nessa linha de pensamento, para realizar tal processo é necessário informar qual é o meu bucket de destino. Em virtude disso, a etapa de mapeamento dos dados ganha relevância pois se não tivesse catalogado toda a nossa base de dados não conseguiríamos realizar as transformações que serão apresentadas a seguir.

#### 3.5.1 REMOÇÃO DE VALORES NULOS

Alguns campos vieram com valores nulos, para poder remover tais valores foi necessário utilizar uma transformação chamada DropNullFields, com isso criamos o nosso job, ou seja a tarefa que irá remover dados nulo da tabela. Com isso, informamos a nossa origem que é o bucket dos arquivos extraídos 20122021, que foram previamente catalogados pelo nosso crawler e logo em seguida aplicamos a transformação desejada no nosso dado e encaminhamos ela ao banco de dados Redshift. A seguir na figura 10 será ilustrado tal processo.

Figura 10 - Remover valores nulo



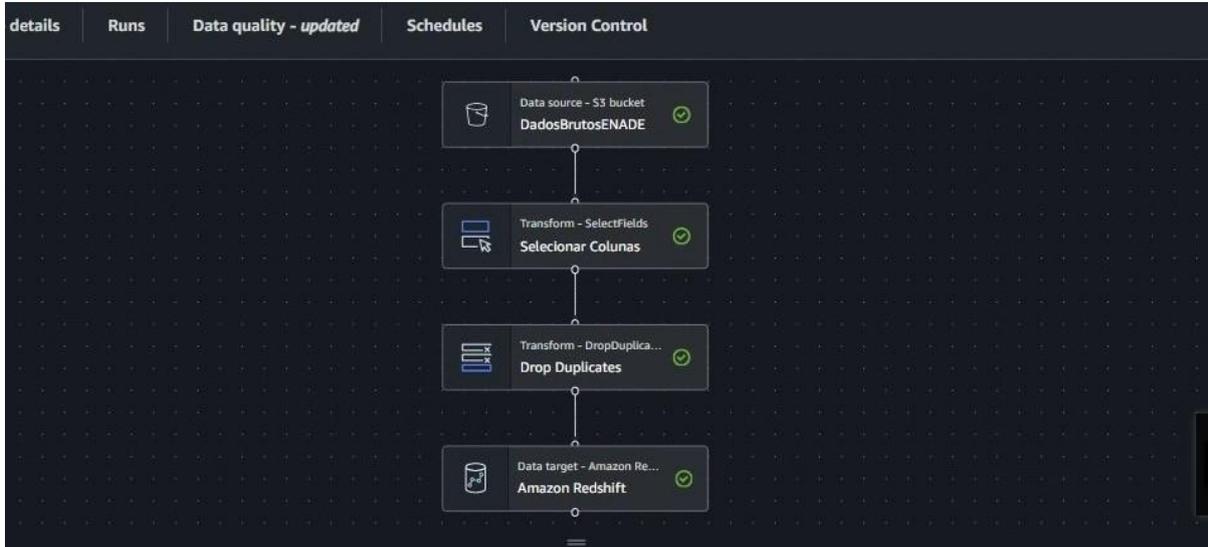
930

Fonte: Próprio Autor (2024)

### 3.5.2 SELEÇÃO DE COLUNAS DESEJADAS E REMOÇÃO DE VALORES DUPLICADOS

Seguindo o padrão da modelagem realizada no capítulo 4.1, foi necessário separar algumas colunas para serem adotadas em determinadas tabelas, com isso a figura 11, ilustra esse processo, além de mostrar que foi necessário eliminar valores duplicados na tabela seja pelo fato de os dados duplicados serem irrelevantes e também pelo fato de que o consumo de memória e o tempo de execução das consultas demandarem um tempo maior para serem realizadas.

Figura 11 - Seleção de Colunas e remoção de valores duplicados



931

Fonte: Próprio Autor (2024)

### 3.5.3 JUNÇÃO DE TABELAS

Uma dificuldade enfrentada durante a realização do projeto foi que alguns campos não foram catalogados pelo Crawler, para resolver esse problema foi necessário criar um crawler por arquivo e não um crawler global que pegasse todos os arquivos txt do meu bucket, nesse sentido a figura 12 ilustra tal processo demonstrando que foi necessário criar um crawler para cada arquivo que eu desejo, criando assim um catálogo para esses arquivos específicos.

Figura 12 - Crawler por arquivo

The screenshot shows the 'Crawlers' page in the AWS Glue console. It includes a search bar, a table of crawler details, and a 'Create crawler' button. The table lists four crawlers, all in a 'Ready' state and successfully completed their last run on April 24, 2024.

Name	State	Schedule	Last run	Last run times...	Log	Table change...
CatalogarBolsista	Ready		Succeeded	April 24, 2024 at ...	View log	1 created
CatalogarCota	Ready		Succeeded	April 24, 2024 at ...	View log	1 created
CatalogarIdade	Ready		Succeeded	April 24, 2024 at ...	View log	1 created
CatalogarSexo	Ready		Succeeded	April 24, 2024 at ...	View log	1 created

Fonte: Próprio Autor (2024)

Com isso foi criado 4 tabelas, dessa forma foi possível realizar a junção pois foi criado uma origem para cada tabela e com isso aplicamos a transformação de Join, para poder juntar as tabelas pelo código, depois selecionamos as colunas desejadas eliminamos valores duplicados e encaminhamos essa nova tabela para o banco com o nome Dim\_Estudante, como demonstrado na figura 13.

932

**Figura 13 - Junção de Tabelas**

Fonte: Próprio Autor (2024)

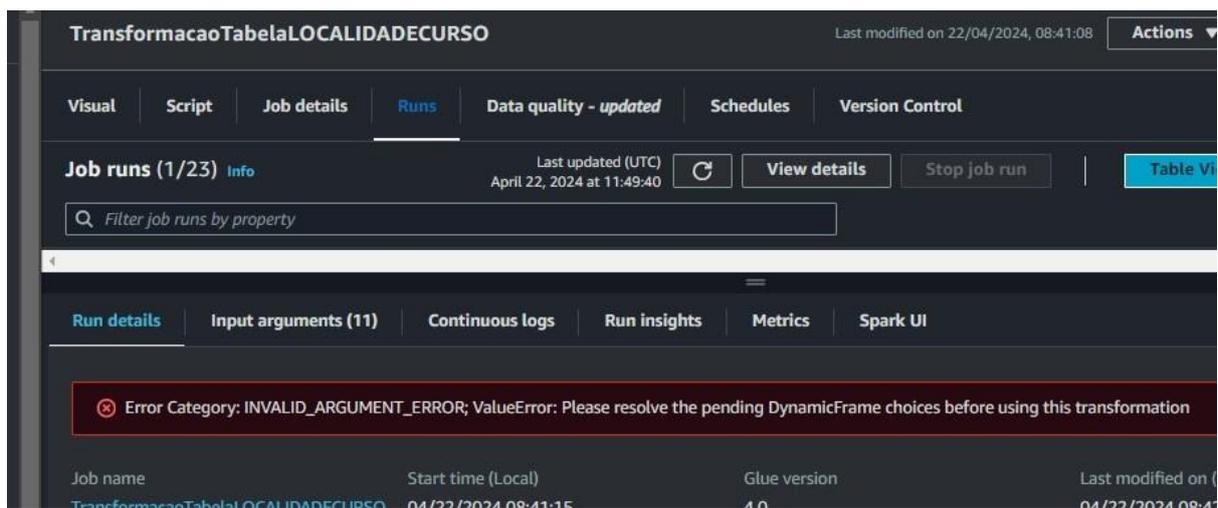
### 3.5.4 TRATANDO DADOS TIPO SUPER

Um dos principais problemas enfrentados durante a elaboração do projeto foi ter que tratar os dados do tipo SUPER, da AWS toda vez que era executado o meu Job o erro DynamicFrame choice era acionado conforme vislumbrado na figura 14.

Nesse sentido, em conformidade com a documentação da AWS, os tipos DynamicFrame, são utilizados quando necessitamos criar uma coluna sem se preocupar com seu schema, ao se utilizar essa metodologia proposta pela AWS, os tipos de dados são definidos no momento de sua invocação. Entretanto, ao se utilizar uma tipagem DynamicFrame ou também chamada struct, acaba por ocasionar múltiplos tipos de dados para uma coluna, ou seja teríamos um array de tipos para uma mesma coluna, desse modo o nosso banco não estava conseguindo interpretar essa multiplicidade de tipos para um mesmo campo, para resolver tal problema foi

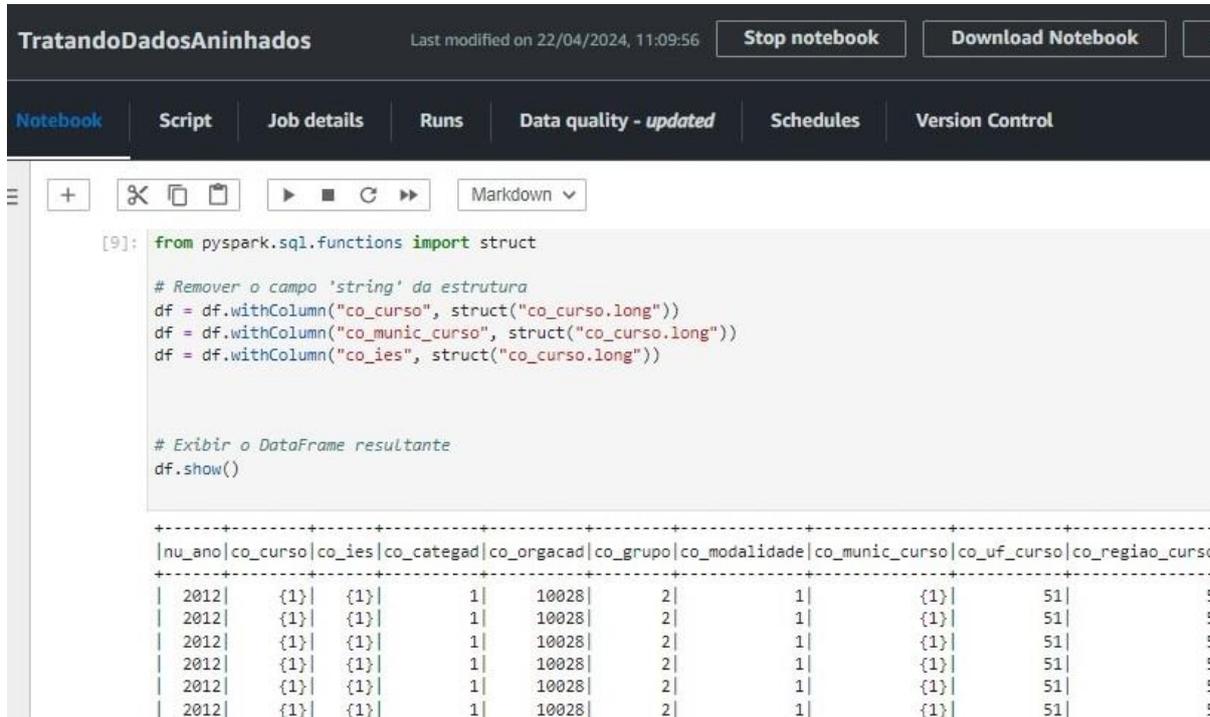
necessário utilizar um serviço da AWS que integra o Jupyter Notebook, conforme apresentado na figura 15, o comando em python remove um dos tipos dos dados, no caso o tipo String que está vazio, ou seja NULL, portanto, ao remover esses dados, a nossa tabela não é afetada, pois não tinha valores atribuídos para esse tipo, após esse tratamento os dados foram encaminhados ao banco.

Figura 14 - Erro Dynamic Frame



Fonte: Próprio Autor (2024)

Figura 15 - Tratando Dados Super



Fonte: Próprio Autor (2024)

### 3.5.5 LINKED SERVER

Quando se está iniciando o aprendizado na AWS, uma das principais limitações é gerenciar os serviços utilizados e monitorar o tempo de utilização dos recursos e consultas SQL, já que a plataforma cobra por esses serviços, como destacado na Figura 16. Para evitar a perda das transformações e dos processos realizados anteriormente, foi necessário estabelecer uma conexão entre o Servidor Remoto e um servidor local, o SQL Server.

Nesse contexto, foi empregada uma funcionalidade do SQL Server chamada linked server, a qual, segundo Garetti (2013), permite a conexão entre diversos servidores, exigindo apenas um driver ODBC do banco de dados RDS. A Figura 17, ilustra o procedimento de conexão do servidor vinculado com o SQL Server.

Portanto, ao utilizar essa funcionalidade do SQL Server, garante que não será necessário reiniciar o trabalho realizado e que algumas consultas podem ser executadas localmente, resultando em uma economia no consumo de consultas e transformações SQL mais simples.

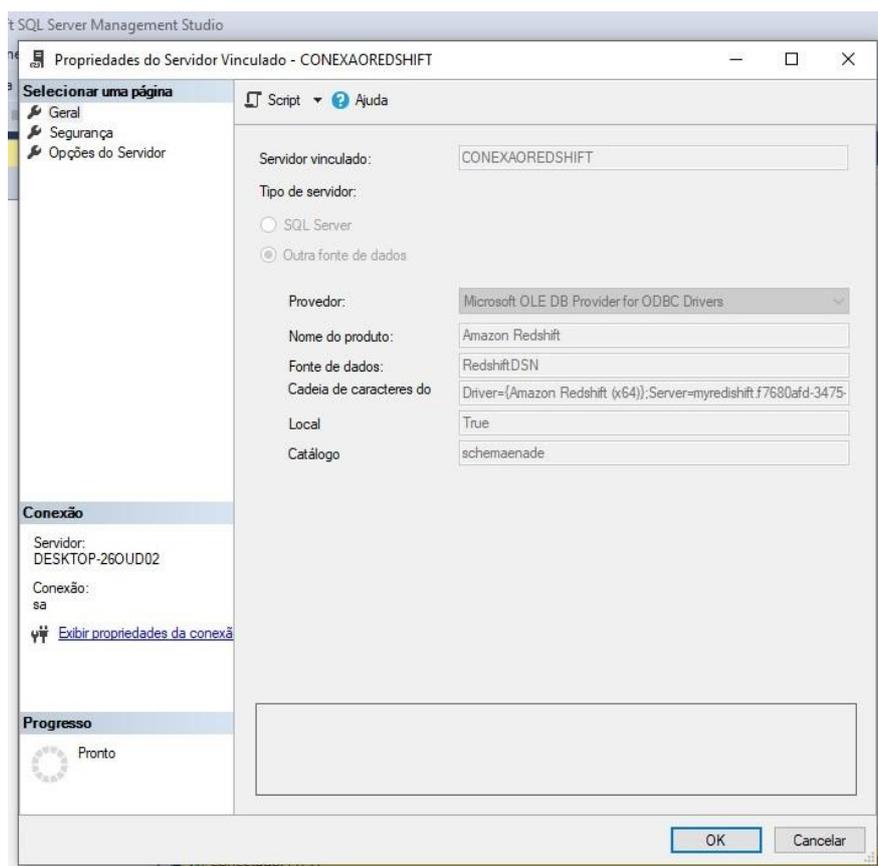
**Figura 16 - Demonstrando do consumo dos laboratórios**

Iniciar os laboratórios de aprendizagem da AWS Academy



Fonte: Próprio Autor (2024)

**Figura 17 - Conexão com o Servidor Vinculado**

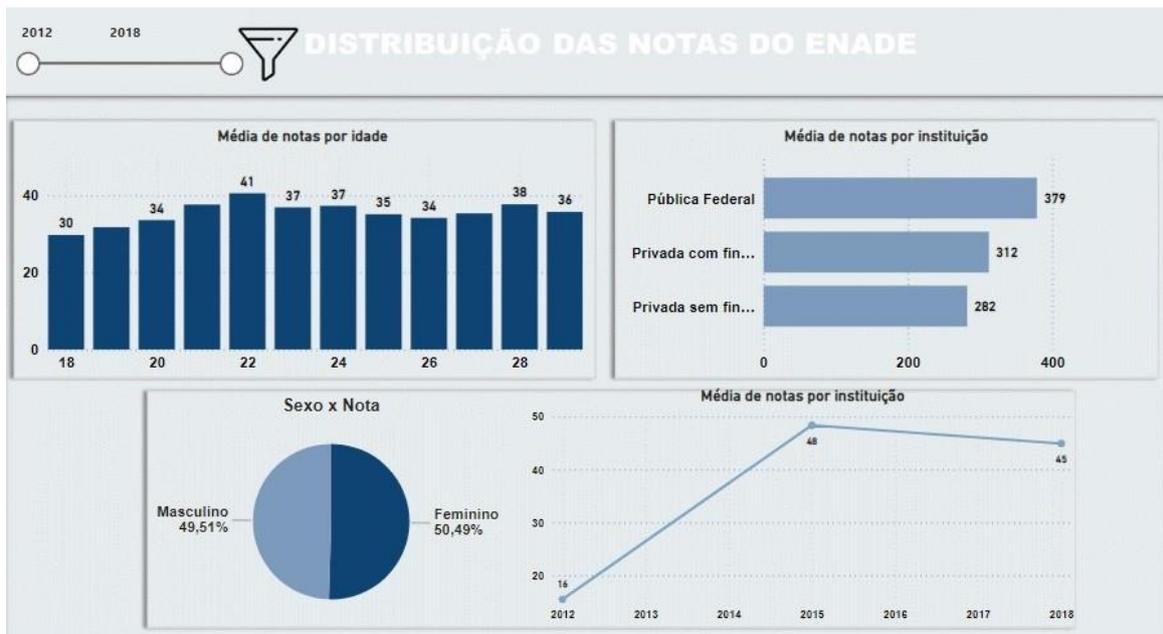


Fonte: Próprio Autor (2024)

### 3.6 DASHBOARD

Após o tratamento e cruzamento de todos os dados, o resultado será um Dashboard, conforme visualizado na figura 18. Este painel contém quatro gráficos distintos. O primeiro gráfico ilustra a média das notas por faixa etária em um determinado período. O segundo gráfico apresenta quais instituições têm as maiores notas obtidas ao longo do tempo. Abaixo desses dois gráficos, há mais dois. Um deles mostra a relação entre sexo e nota, permitindo identificar qual gênero está obtendo as maiores notas ao longo dos anos. O último gráfico exibe a evolução das médias de notas obtidas ao longo do tempo. Além disso, possui um filtro por anos, possibilitando a verificação de tendências em diferentes intervalos temporais. Esse dashboard proporciona uma visão abrangente e detalhada do desempenho acadêmico, facilitando a análise de tendências e comparações entre diferentes categorias ao longo do tempo.

Figura18-Dashboard desenvolvido



Fonte: Próprio Autor (2024)

#### 4 CONSIDERAÇÕES FINAIS

Ao final desta monografia, percebe-se a importância de obter dados tratados e cruzados entre si para serem apresentados de forma dinâmica em um dashboard. Nessa linha de pensamento, é evidente a relevância desses dados, pois, com eles em mãos, é possível realizar uma análise social do desempenho dos alunos por idade, sexo e instituições. Com essas informações, pode-se traçar planos para melhorar o ensino nas instituições.

Nesse contexto, diversos objetivos foram alcançados, como a extração de dados do governo, a criação de uma rotina de automação para extração de arquivos e a visualização de dados. No entanto, alguns objetivos não foram atingidos e podem se tornar temas para futuros estudos. Entre esses objetivos estão: identificar as regiões com as maiores notas, determinar os cursos com as melhores avaliações e localizar onde esses cursos mais bem ranqueados estão distribuídos no Brasil. Por fim, não foi possível traçar os dados de cota e bolsas de estudos sendo assim esses os limites do estudo apresentado.

Considero de suma importância que um estudante tenha um projeto que possa chamar de seu. Ao final deste projeto, sinto-me realizado, pois estive em contato com uma área pela qual me interessei desde o segundo semestre da faculdade. Através da iniciação científica, pude me aprofundar nos principais desafios da área de ciência de dados e desenvolver muitas habilidades que, tenho certeza, serão relevantes na carreira que pretendo seguir.

#### 5 REFERÊNCIAS BIBLIOGRÁFICAS

ARRAIS, Karolayne. **Construção de um pipeline de dados utilizando serviços da nuvem**. Trabalho de Conclusão de Curso (Bacharelado em Engenharia de Computação). Universidade Federal de São Carlos Centro de Ciência Exatas e de Tecnologia, São Carlos. P. 91. 2022.

AWS. **AWS**. Plataforma de Documentação da plataforma utilizada durante o projeto. Disponível em: [https://docs.aws.amazon.com/pt\\_br/glue/latest/dg/aws-glue-api-crawler-pyspark-extensions-dynamic-frame.html](https://docs.aws.amazon.com/pt_br/glue/latest/dg/aws-glue-api-crawler-pyspark-extensions-dynamic-frame.html) . Acesso em: 25 maio.2024.

BELLO, Luis. BRITO, Vinícius. Uma em cada quatro mulheres de 15 a 29 anos não estudava e nem estava ocupada em 2023. **Agência IBGE Notícias**, 2024. Disponível em: <https://agenciadenoticias.ibge.gov.br/agencia-noticias/2012-agencia-de-noticias/noticias/39531-uma-em-cada-quatro-mulheres-de-15-a-29-anos-nao-estudava-e-nem-estava-ocupada-em-2023#:~:text=Em%202023%2C%20a%20taxa%20de,atrasados%2C%20frequentando%20a%20educa%C3%A7%C3%A3o%20b%C3%A1sica> .Acesso em: 25 maio. 2024.

EMPRESA PARCEIRA. **Saraiva Educação**, 2022. Plataforma para a divulgação de conhecimento. Disponível em: <https://blog.saraivaeducacao.com.br/enade/>. Acesso em: 25 maio. 2024.

FERNANDES, Naira. **Uso de Data Lake na geração de informações de mercado para o ramo estético do Brasil: Um estudo de caso**. 2022. 49. Projeto de Pesquisa .Instituto Federal de Santa Catarina, Caçador, Santa Catarina, 2022.

GARETTI, Luiz. Linked Server pra que serve e quando utilizar ?. **We are developers**, 2013. Disponível em: <https://imasters.com.br/data/linked-server-pra-que-serve-e-quando-utilizar>. Acesso em: 25 maio. 2024.

HAUGG, Igor. **Estudo dos Tipos de Serviços de Computação em Nuvem Desde uma Perspectiva de Integração de Aplicações**. 2014. 61. Trabalho de conclusão de Curso. Universidade Regional do Noroeste do Estado do Rio Grande do Sul – UNIJUI Grupo de Pesquisa em Computação Aplicada (GCA). 2014.

IAAS, PAAS, SAAS: ENTENDA OS MODELOS DE NUVENS E SUAS FINALIDADES. **Noventiq**, 2022. Disponível em: <https://noventiq.com.br/sobre-a-empresa/blog/iaas-paas-saas-nuvem>.

MINISTÉRIO DA EDUCAÇÃO. **Gov.br**, 2022. Plataforma responsável por divulgar dados e análises sobre a educação no Brasil. Disponível em: <http://portal.mec.gov.br/component/tags/tag/32044-censo-da-educacao-superior> .Acesso em: 25 maio. 2024.

NUNES, Leidiane. **Ciência de Dados como método de transformação de dados em informação**. Trabalho de Conclusão de Curso (Tecnólogo em Sistemas Para Internet.). Instituto Federal de Educação, Ciência e Tecnologia do Sertão Pernambucano, Pernambuco, p. 24. 2022.

PEREIRA, Hélder, SOUZA, José, SCHULZE, Bruno, MURY, Antonio. **Computação Em Novembro**. 2011. Maranhão, São Luís, Brazil.

SOUZA, Amaranta; SALAZAR, Eder; SOARES, Liriane; BRUSCHI, Paula; ALMEIDA, Tayná; RIBEIRO, Verônica. **Ciência de Dados, Business Intelligence e Big Data: conceitos e aplicações**. 1 ed. Curitiba: Appris, 2021.

---

SYOZI, Ricardo. O que é a AWS [Amazon Web Services]. **Tecnoblog**, 2021. Disponível em: <https://tecnoblog.net/responde/o-que-e-a-aws-amazon-web-services/>. Acesso em: 06 maio. 2024.