GUSTAVO LIRA GIRARDI DE GÓES

ESTRUTURAÇÃO DE BANCO DE DADOS NÃO-RELACIONAL PARA AS INFORMAÇÕES DE POÇOS DE PETRÓLEO DAS BACIAS SEDIMENTARES CONTINENTAIS BRASILEIRAS

Trabalho de Conclusão de Curso apresentado à Escola Politécnica da Universidade de São Paulo para obtenção do diploma de Engenharia de Petróleo.

SANTOS

2021

GUSTAVO LIRA GIRARDI DE GÓES

ESTRUTURAÇÃO DE BANCO DE DADOS NÃO-RELACIONAL PARA AS INFORMAÇÕES DE POÇOS DE PETRÓLEO DAS BACIAS SEDIMENTARES CONTINENTAIS BRASILEIRAS

Trabalho de Conclusão de Curso apresentado à Escola Politécnica da Universidade de São Paulo para obtenção do diploma de Engenharia de Petróleo.

Área de concentração: Análise de dados

Orientador: Prof. Dr. Cleyton de Carvalho Carneiro

Coorientador: Rodrigo César de Teixeira Gouvêa

SANTOS

2021

FICHA CATALOGRÁFICA

de Góes, Gustavo Lira Girardi
ESTRUTURAÇÃO DE BANCO DE DADOS NÃO-RELACIONAL PARA AS INFORMAÇÕES DE POÇOS DE PETRÓLEO DAS BACIAS SEDIMENTARES CONTINENTAIS BRASILEIRAS / G. L. G. de Góes -- São Paulo, 2021.
41 p.

Trabalho de Formatura - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Minas e de Petróleo.

1.Mineração de dados 2.Base de dados não-relacional 3.Ciência de dados 4.Bacias onshore 5.Catalogação de poços I.Universidade de São Paulo.
Escola Politécnica. Departamento de Engenharia de Minas e de Petróleo II.t.

AGRADECIMENTOS

Acima de tudo gostaria de agradecer à minha mãe, Dairce da Silva Lira. Ela, que abdicou de uma série de oportunidades e vivências durante sua vida para me criar sempre da melhor forma possível, é a pedra fundamental do homem que me tornei e sem ela e toda a minha família eu não chegaria até aqui. Ao Prof. Dr. Cleyton, que acreditou em meu trabalho de conclusão e ao Rodrigo, que nos trouxe os ensinamentos necessários e me suportou com a programação em Python, agradeço a oportunidade e confiança. Agradeço também à minha companheira Natália, que se manteve ao meu lado na maior parte deste longo trajeto que foi minha graduação e ainda contribuiu com muitos ensinamentos para me possibilitar a execução desse trabalho.

Sou grato também à oportunidade de ter fundado o Centro Acadêmico da Poli Santos (CAPS) e a CORE Jr (empresa júnior dos politécnicos caiçara), que enriqueceram e muito a minha experiência acadêmica e aos amigos que ganhei após minha estadia em Santos, sejam eles alunos da graduação ou funcionários da Poli Santos.

RESUMO

O acesso e a visualização de informação de dados de poços de exploração e produção de petróleo consistem em etapas fundamentais para o desenvolvimento de projetos diversos. Em bases de dados amplas, no entanto, essas etapas demandam extenso investimento de tempo. Isso ocorre porque as ferramentas de busca mais usuais nem sempre acessam o conteúdo dos arquivos de poço, onde constam todos os registros relativos ao conteúdo dos perfis, profundidades e rodadas de aquisição. A Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP) disponibilizou em mídia física no segundo trimestre de 2021 todos os dados de exploração e produção relativos às bacias *onshore* brasileiras. Dentro desse conteúdo estão os dados de perfilagem de poços. Esses dados compõem um acervo relevante, com metadados diversos dotados de uma variedade de perfis adquiridos por diferentes companhias. Cada poço, portanto, possui particularidades relativas ao conteúdo de perfilagem, profundidade, período produtivo, distribuição temporal, disponibilidade de perfil composto, dentre outros. Este trabalho visa, por meio de ferramentas computacionais, realizar a análise exploratória inicial dos dados e desenvolver um sistema indexador para triagem das informações de cada poço, permitindo a observação e análise prévia dos dados do conteúdo disponibilizado, e também dos dados de perfuração com os perfis disponíveis para mais de 21 mil poços por meio de uma base de dados não relacional. Com mais de 14 mil poços processados com arquivos do tipo LIS e DLIS, é possível fazer uma análise e triagem desses dados via *dashboard* e utilizá-los no processo de localização e catalogação de informações específicas em trabalhos que se utilizarão da base de dados.

Palavras-chave: mineração de dados, base de dados não-relacional, ciência de dados, bacias *onshore*, catalogação de poços.

ABSTRACT

Accessing and visualizing information from oil exploration and production wells are fundamental steps for the development of several projects. In large databases, however, these steps require an extensive time investment since the most common search tools do not always access the content of the well files, which contain all the records related to the content of the logs, depths and acquisition rounds. In the second quarter of 2021, the National Agency for Petroleum, Natural Gas and Biofuels (ANP) made available, in physical media, all exploration and production data related to the Brazilian onshore basins. Within those data, it is possible to find well logging information. These data make up a relevant collection, with different metadata endowed with a variety of logs acquired by different companies. Therefore, each well has particularities related to logging content, depth, production period, temporal distribution, availability of composite log, among others. This work aims, through computational tools, to carry out the initial exploratory analysis of the data and develop an indexing system for screening the information from each well, allowing the observation and prior analysis of the data available, as well as the drilling data with the logs available for more than 21,000 wells through a non-relational database. With more than 14,000 wells processed with LIS and DLIS files, it is possible to analyze and filter this data via dashboards and use them for locating and cataloging specific information in works that will use the database.

Keywords: data mining, non-relational database, data science, onshore basins, well cataloging.

LISTA DE FIGURAS

LISTA DE TABELAS

SUMÁRIO

1 INTRODUÇÃO

No início de 2021, a Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP) disponibilizou em mídia física diversos dados de exploração e produção relativos às bacias *onshore* brasileiras. Dentro desse conteúdo estão dados de poços que compõe um acervo relevante, com metadados diversos dotados de uma variedade de poços adquiridos por diversas companhias.

O conteúdo da base de dados disponibilizada engloba 21.307 poços distribuídos em 23 bacias geológicas de 22 estados brasileiros. Esses dados compactados resultam em mais de 1,3 TB de informação não integrada.

Um dos elementos básicos e comum à grande parte dos poços disponibilizados é o Arquivo Geral de Poço (AGP), que são extratos digitais com diversas informações dos poços cedidos pelas operadoras à ANP em formato .TXT e era recorrente para os poços brasileiros perfurados até 2009 (vide figura 1).

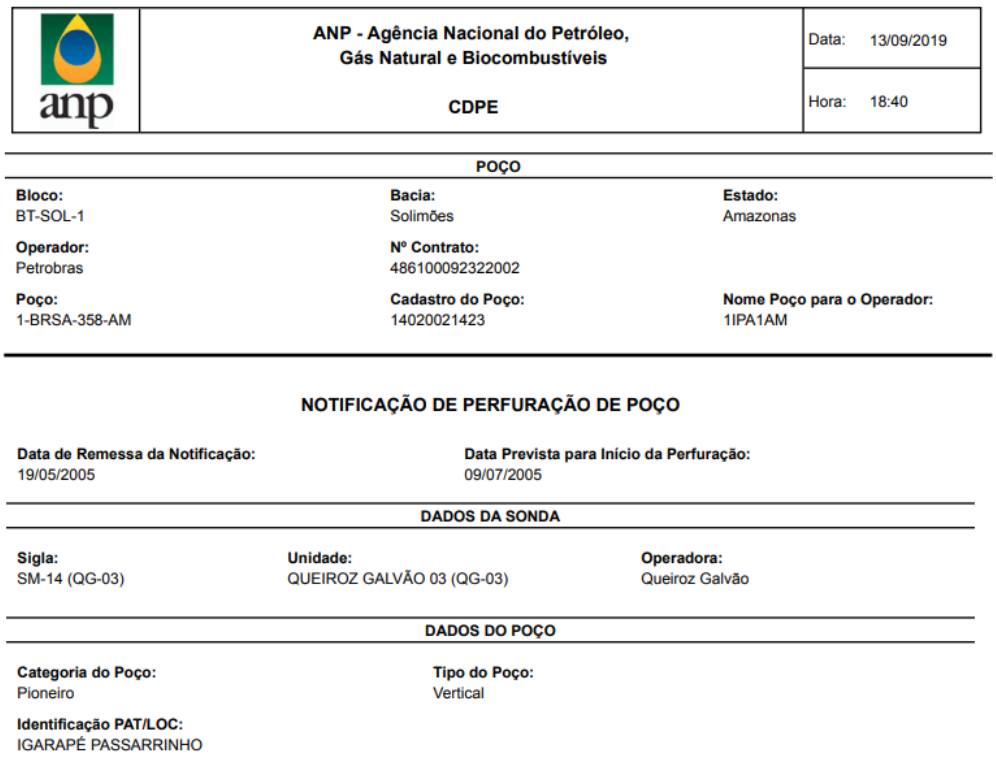
```
POCO           :   1AJ  0001  AM

IDENTIFICADOR  :   SB19F43SO84C83B                    BACIA          :   SOLIMOES
CAMPO          :                                      DATUM          :   SAD-69
LATITUDE       :    -4.97711 (   -4  58 37.6)         LONGITUDE      :   -66.44411 (  -66  26 38.8)
MERID.CENTRAL  :   - 69                               COORD.UTM FUNDO:    9449310.6S  783445.6W
DISTRITO       :   E&P-AM                             ESTADO         :   AM
MUNICIPIO      :   CARAUARI                           CODIGO POCO    :   007442
QUADRICULA     :   34944978                           MESA ROTATIVA  :     81.0
B.A.P          :      75.0                            P.F.SONDADOR   :   3340.0
PROF.MX.PERF.  :   3337.0                             METROS PERF.   :   3340.0
MAIOR PROF.    :   3340.0                             INICIO         :   25/01/82
TERMINO        :   26/04/82                           DATA LIB.SONDA :   29/04/82
ULT.RECLASSIF. :   06/05/82                           ULT.RECLASSIF. :   LOCACAO
SITUACAO LOC.  :   CONCLUIDO                          DATA APROV.LOC :   22/09/81
CADASTRO       :   00457                              SEDOC          :   0268
SONDA          :   SM-2                               FORMACAO P.F.  :   F JANDIA
TERRA / MAR    :   TERRA                              ACS RISCO      :
CONTR. RISCO   :                                      DOC.APROV.LOC. :   DEPEX-T-52.412/1981
NOME           :   ARUAJA.1                           CLASSIFICACAO  :   PIONEIRO
RECLASSIFICACAO:   SECO SEM INDICACAO DE H.C.
MC C.UTM BASE  :   - 69                               COORD.UTM BASE :     9449310.6S    783445.6W
DADOS  LOCACAO :
MC LOCACAO     :   - 69                               COORD.UTM LOC. :     9449310.5S   783445.6W
DATUM LOCACAO  :   SAD-69                             EMPR. OPER.1/2 :    PARKER
DIRECIONAL?    :   VERTICAL
HISTORICO DA RECLASSIFICACAO -      06/05/82   -  SECO SEM INDICACAO DE H.C.                          - RECLASSIFICACAO
----------------
REVESTIMENTOS -         TIPO           PROFUNDIDADE        DIAMETRO       IND. DE RECUPERACAO
----------------
                   TUBO CONDUTOR        10.5  (   70.5)    20               NAO
                   REV. SUPERFICIE     817.0  ( -736.0)    13 3/8           NAO
                   REV. INTERMED.     1935.0  (-1854.0)     9 5/8           NAO
```

Figura 1 - Arquivo AGP do poço 1-AJ-AM, da bacia do Solimões

No início dos anos 2010, com a expansão do número de empresas operando nas bacias de propriedade do Estado foi proposto e aplicado o Padrão ANP10, que é uma coletânea de relatórios cedidos pelas empresas operadoras das áreas contratadas e regulamentado pela ANP.

O padrão ANP10 surgiu à partir da Nota Técnica nº 073/2016/SDT (ANP, 2016), que passou a regulamentar e exigir um envio padronizado, criando a Consolidação de Dados de Poço Exploratório (CDPE). Os CDPE (figura 2) são relatórios com metadados sobre a evolução do poço submetidos em até um ano após a data de conclusão do poço e possuem um layout padrão digitalizado, essencialmente em arquivos pdf, com a seguinte parametrização:

- Notificação de Perfuração de Poço (NPP)
- Relatório Final de Completação de Poço (RFCP)
- Relatório Final de Poço Explota(ra)tório (RFP)
- Relatório Final de Abandono de Poço (RFAP)
- Notificação de Conclusão de Reentrada em poço (NCRP)
- Notificação de Perfilagem Realizada (NPR)



ANP - Agência Nacional do Petróleo, Gás Natural e Biocombustíveis

CDPE

Data: 13/09/2019

Hora: 18:40

POÇO

Bloco: BT-SOL-1	**Bacia:** Solimões	**Estado:** Amazonas
Operador: Petrobras	**Nº Contrato:** 486100092322002	
Poço: 1-BRSA-358-AM	**Cadastro do Poço:** 14020021423	**Nome Poço para o Operador:** 1IPA1AM

NOTIFICAÇÃO DE PERFURAÇÃO DE POÇO

Data de Remessa da Notificação: 19/05/2005	**Data Prevista para Início da Perfuração:** 09/07/2005

DADOS DA SONDA

Sigla: SM-14 (QG-03)	**Unidade:** QUEIROZ GALVÃO 03 (QG-03)	**Operadora:** Queiroz Galvão

DADOS DO POÇO

Categoria do Poço: Pioneiro	**Tipo do Poço:** Vertical
Identificação PAT/LOC: IGARAPÉ PASSARRINHO	

Figura 2 - Exemplo de relatório CDPE do poço 1-BRSA-358-AM, da bacia do Solimões

Outro pilar fundamental desse projeto de mineração dos dados cedidos pela ANP é a extensão DLIS (*Digital Log Interchange Standard*). Os arquivos .DLIS proporcionam um compilado de dados de poços extraídos a partir das ferramentas de perfilagem como porosidade, raios gama, resistividade, ressonância e diâmetro do poço. Foi inicialmente criado em 1991 e até hoje é utilizado como uma poderosa ferramenta de padronização de dados e armazenamento de informações sobre poços de petróleo, uma vez que facilita a utilização em diversos ambientes de estudo e pesquisa.

Ao extrair as principais informações dos arquivos DLIS permitimos ao usuário a prévia compreensão e acesso às curvas dos perfis geofísicos sem a necessidade de acessar as pastas e subpastas de arquivos individualmente (figura 3).

```
[60]:
origin, *origin_tail = f.origins
origin.describe()
```

```
[60]: ------
Origin
------
name   : DLIS_DEFINING_ORIGIN
origin : 94
copy   : 0

Logical file ID          : HALS_DSI_TLD_MCFL_026PUC
File set name and number : PETRO/1-IPA-1_R2 / 41
File number and type     : 31 / PLAYBACK

Field                    : CENTRO-NORTE DO BT-SOL1
Well (id/name)           : 14020021423 / 1-BRSA-358-AM
Produced by (code/name)  : 440 / Schlumberger
Produced for             : PETROBRAS
Order number             : BRUR 0805_001
Run number               : 2
Descent number           : -1
Created                  : 2005-08-11 17:23:37

Created by               : OP, (version: 13C0-300)
Other programs/services  : HALS-B: HILT Azimuthal Laterolog Sonde B
                           DSST-B: Dipole Shear Imager - B
                           HILTD: High resolution Integrated Logging Tool-DTS
                           DTCH: DTS Telemetry Tool
                           BSP: Bridle SP
                           LEHQT: Logging Equipment Head - QT
```

Figura 3 - Exemplo de obtenção de dados .DLIS do poço 1-BRSA-358-AM, da bacia do Solimões

O grande obstáculo para análises comparativas mais detalhadas a partir da base de dados disponibilizada pela ANP se relaciona às buscas massivas eficientes. Essas buscas, atualmente, disponibilizam informações sobre poços brasileiros *onshore* em inúmeras pastas e arquivos avulsos. Isso dificulta também a possibilidade de visualização conjunta dos dados e a agregação dos mesmos de forma padronizada para *machine learning*, por exemplo. Há ainda diferentes tipos de arquivos que caracterizam um poço, evidenciando a necessidade de saber, por exemplo, o que cada pasta de um poço específico contém. Atualmente, da forma que essas informações são apresentadas, é necessária uma significativa demanda de recursos para saber quais poços possuem determinados perfis, análises, arquivos e outros aspectos característicos relevantes.

1.1 Objetivos

Este trabalho tem como objetivo geral proporcionar a exploração assistida e a vizualização dinâmica de informações de poços de petróleo *onshore* brasileiros a partir de uma base de dados integrada. A nova base de dados se baseia em informações do arquivo geral do poço - AGP (.TXT), bem como nos canais que compõem os arquivos *digital log interchange standard* (.DLIS) e *log interchange standard* (.LIS).

Como objetivos específicos, a pesquisa visa:

- Desenvolver uma base de dados não relacional composta por informações extraídas dos arquivos AGP, .LIS e .DLIS de todas as bacias continentais brasileiras disponibilizadas pela ANP em 2021;
- Desenvolver indexador para varredura e triagem das informações contidas nos poços, trazendo os metadados provenientes dos frames e canais dos arquivos digitais de perfilagem;
- Associar a base de dados a um formato de visualização dinâmica do conteúdo dos perfis, facilitando a utilização por usuários de múltiplas áreas;
- Desenvolver uma plataforma *user friendly* (amigável) de consulta aos metadados disponíveis para cada poço *onshore* disponibilizado pela ANP,

permitindo uma fácil expansão para - por exemplo - futuramente agregar o material também *offshore*.

1.2 Justificativa

O início da produção de petróleo em terras brasileiras remete ao começo do século XX, no interior do Estado de São Paulo. Porém, somente na década de 1920 tivemos os primeiros registros de dados oficiais de exploração. No final dos anos 1930 foram descobertos os primeiros campos produtores *onshore* na Bacia do Recôncavo. Com a criação da Petróleo Brasileiro S.A. (Petrobras) em 1953, iniciou-se um grande processo de exploração e produção de petróleo *onshore* em bacias como a do Solimões, Recôncavo, Alagoas e Sergipe.

Com a quebra do monopólio do petróleo pela Petrobras em 1997 foi possível observar cada vez mais empresas de pequeno e médio porte empreendendo na exploração de poços terrestres, já que as gigantes multinacionais focaram em campos ultraprofundos e de grande porte. Dada a variedade de empresas e também bacias, campos e principalmente poços, foi criada a agência regulamentadora do setor, a ANP (Agência Nacional do Petróleo, Gás Natural e Biocombustíveis).

A partir da evolução das ferramentas de gestão e padronização dos dados, a ANP passou a organizar e estruturar as suas bases de dados de poços por todo o país. Nesse novo formato, as bases foram divididas comumente em cinco categorias, onde as duas primeiras serão melhor abordadas ao longo desse trabalho:

- Arquivo Geral de Poços
- Pastas de Poços
- Dados geoquímicos
- Perfis Compostos
- Curvas de Perfis de Poço

Em 2021 a ANP deu acesso amplo e irrestrito à base de dados das bacias sedimentares continentais brasileiras. Esta base compreende um amplo volume de informações, incluindo os dados de perfis de poços. No entanto, as ferramentas de

exploração de bases de dados da própria agência possuem recursos limitados de visualização para finalidades diversas, bem como um processamento lento. Como exemplo, em uma simples busca por perfis de poços que tenham informações de raios gama e densidade, situados na bacia do Solimões, seria necessário analisar o conteúdo de 2.014 pastas e subpastas com 3.891 arquivos diferentes distribuídos em mais de 82 GB de dados.

A complexidade associada a compreensão das informações de poços é o objeto central deste projeto. Apesar de cada pasta de poço da base de dados disponibilizada originalmente conter os respectivos arquivos de extensão .DLIS, não é possível acessar ao conteúdo dos canais de perfilagem associado a cada poço sem a buscar item a item, pasta por pasta. Entende-se, portanto, a importância da compilação desses metadados em uma base indexada separada, que possa ser consultada de forma rápida e intuitiva para a aquisição e compilação dessas informações.

A estruturação do banco de metadados dessa rede de aproximadamente 21 mil poços de bacias terrestres, portanto, feita em uma base de dados não relacional, permite a compilação de dados semiestruturados de forma a possibilitar sua fácil concatenação e futuras buscas (*queries*), com fácil acesso a estatísticas e informações sobre uma rede de milhares de poços *onshore* brasileiros. Esse processo permite que empresas interessadas em investir em campos onshore brasileiros, assim como pesquisadores e estudantes tenham a possibilidade de analisar diferentes características e informações de toda essa estrutura de metadados para fins acadêmicos, educacionais, modelamento e ciência de dados , além de embasar a seleção de amostragem para treinamentos de *machine learning,* aprendizado profundo, entre outros.

2 REVISÃO BIBLIOGRÁFICA

A crescente quantidade de dados disponíveis nas diversas áreas da ciência, aliados a expansão do leque das aplicações de ciência de dados, tem permitido melhoras consideráveis do entendimento de processos e do modelamento de sistemas complexos. Isso foi possível e impulsionado principalmente pelo aumento da capacidade de processamento dos computadores modernos e pela computação em nuvem.

2.1 Banco de Dados

A Ciência de Dados, de forma resumida, é responsável pela aquisição/coleção, avaliação e processamento de dados para a extração de informações. Portanto, o fluxo de trabalho de um cientista de dados se inicia pelo Banco de Dados, que pode ser visto como uma coleção organizada na qual se armazenam os dados. Contudo, bancos de dados técnicos voltados para aplicativos científicos ou de engenharia muitas vezes fornecem grandes desafios para o desenvolvedor, sendo necessário que se tenha um bom entendimento dos dados bem como da demanda do usuário final (ANGWIN; NELSM; SYRETT, 1996).

De acordo com Navak et al. (2013), desenvolvimentos mais recentes no campo dos bancos de dados em programação criou uma distinção forte entre bancos de dados relacionais (*Standard Querry Language* ou SQL) e não relacionais (NoSQL), que apresentam diferenças fundamentais com relação a estrutura, armazenamento, relação e principalmente na construção de buscas (*queries*) dos dados.

2.2 Banco de Dados Relacional (SQL)

O conceito de Banco de Dados Relacional foi primeiramente apresentado em CODD, 1970 e desde então dominou o mercado de banco de dados. Esse modelo se baseia em estruturas hierárquicas ou de navegação para organização dos dados em tabelas, cujas informações são distribuídas em linhas e colunas.

Cada tabela contém um ou mais dados em colunas, e cada linha é uma instância exclusiva de dados ou chave para os dados definidos nas colunas. A presença de uma ou mais colunas de chaves primárias e chaves estrangeiras garante o caráter relacional dessa estruturação de dados.

2.3 Banco de Dados Não Relacional *Not Only SQL* (NoSQL)

Devido ao aumento exponencial da quantidade de dados disponíveis, devido ao aumento da acessibilidade à internet e aos dispositivos móveis *smartphones*, gerou-se a necessidade de alternativas ao modelo relacional de banco de dados, o qual não gerência de forma efetiva essa quantidade de dados. Como resultado, surgiram os bancos de dados não relacionais, ou não somente SQL (NoSQL).

Essa estrutura se baseia no modelo BASE (Basically Available, Soft State and Eventually Consistent)(JATIN; BATRA, 2016), de natureza distribuída, permitindo a disponibilidade parcial dos dados mesmo quando partes do banco de dados não estão operacionais ou não podem ser alcançados. Esse modelo garante aos bancos de dados não relacionais escalabilidade, melhor performance e maiores níveis de disponibilidade aos seus usuários, além de utilizar, de forma geral, menor espaço de armazenamento. Essa estrutura também permite o armazenamento de dados de qualquer tipo, como texto, imagem, áudio e vídeos, dentro de um mesmo banco de dados (FRACZEK; PLECHAWSKA-WOJCIK, 2017; PHIRI; KUNDA, 2017).

Existem diversos tipos de bancos não relacionais como Banco de Documentos (utiliza o formato JSON), Bancos de Dados baseados na estrutura Chave-Valor, Armazenamento de dados em grafos (através de vértices e arestas), Armazenamento de Séries Temporais, Armazenamento de Objetos, Armazenamento por Índice Externo, entre outros (“Dados não relacionais e NoSQL - Azure Architecture Center | Microsoft Docs”).

2.4 Preparação, limpeza e pré-processamento

É frequentemente citado que 80% do tempo do processamento de dados é gasto na limpeza e preparação de dados (COX, 2004; ENDEL; PIRINGER, 2015; KANDEL

et al., 2011; LOHR, 2014). Portanto, esse processo de preparação, limpeza e pré-precessamento dos dados é de fundamental importância para qualquer tipo de análise, seja estatística ou como preparação para a aplicação em modelos de *machine learning*. A preparação, limpeza e pré-processamento é contínua, sendo necessária com a coleta de novos dados (WICKHAM, 2014).

O processo de limpeza e preparação dos dados, muitas vezes citado como *Data Munging* ou *Data Wrangling*, é o processo de transformação dos dados da sua forma "bruta" ao formato de interesse para a análise escolhida e a organização desses dados de forma que se possa avançar para os estágios de processamento, em outras palavras, o processo de exploração e transformação de dados de forma iterativa de forma a permitir sua análise (ENDEL; PIRINGER, 2015).

Dentro desse processo, se faz necessário muitas vezes a aplicação de reformatação, extração, correção de outliers, conversão de tipo e mapeamento dos dados. O mapeamento de dados (*Data Mapping*) envolve a identificação dos campos de dados de origem e a sua aplicação nos campos de dados de destino, dentro do processo de estruturação dos dados (KANDEL et al., 2011).

Existem diversos processos que podem ou não ser empregados no processo de estruturação dos dados para processamento, os quais são fundamentalmente dependentes da natureza e origem dos dados a serem trabalhados. Portanto, as etapas de limpeza e pré-processamento de dados são específicas de cada caso. De forma geral, como apontado por ENDEL; PIRINGER, 2015, os desafios são geralmente relacionados a atributos, qualidade, fusão e conexão, reproducibilidade e documentação, incertezas, erro, e transformação e edição de dados.

O conceito de utilidade dos dados emerge nesse contexto, uma vez que o processo de preparação dos dados é fundamentalmente um processo de tornar os dados úteis. Um dado é útil se for usável, verossímil e responsivel à investigação a que ele está sendo aplicado (KANDEL et al., 2011). Dessa forma, o objetivo fundamental do pré-processamento dos dados é criar dados estruturados, unificados e úteis.

2.5 Perfilagem de poços

De acordo com De Andrade (2019), através de ferramentas de perfilagem é feito um registro de propriedades físico-químicas das formações rochosas encontradas em um poço de petróleo, as quais são usadas para traçar um perfil de poço. A perfilagem pode utilizar-se de métodos diretos e indiretos para gerar informações sobre o reservatório e a possibilidade de produção de hidrocarbonetos naquela região. Através desses métodos, obtém-se diversas características do poço ao longo de sua varredura, como resistividade, potencial sônico, caliper e raios gama, auxiliando, por exemplo, a calcular porosidade ou saturação de hidrocarbonetos e identificar eventuais fraturas na formação ou rochas específicas.

2.6 Visualização em dashboard

Dashboards são apresentações visuais de informações relevantes, consolidadas em uma só tela para que todo o material possa ser monitorado de forma ágil e eficiente. Um dos principais softwares para criação e gerenciamento de dashboards é o Power BI, um conjunto de ferramentas criadas pela Microsoft que permite transformar, analisar e visualizar fontes de dados não relacionadas, as quais podem ser oriundas, por exemplo, do Microsoft Excel, Python, Azure, SQL, entre outros. Com este software é possível criar dashboards que proporcionam análises e relações entre fontes de dados independentes, agilizando processos e conectando informações (exemplo na Figura 4). O Power BI dispõe de mais de 130 tipos de visualizações, como gráficos, tabelas e nuvem de palavras, que conectam e analisam seus dados, trazendo *insights* coerentes e relevantes (DATA B INTELIGÊNCIA, 2020).

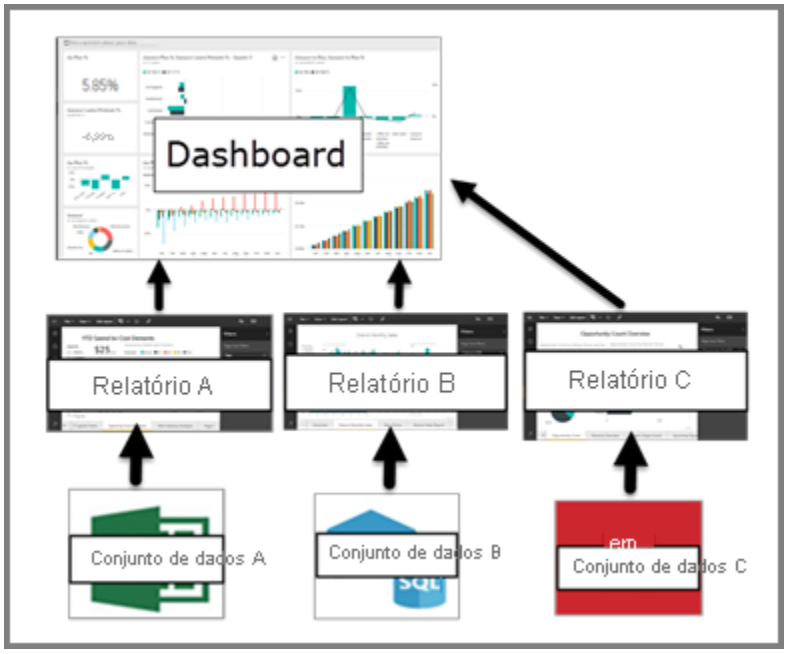


Figura 4 - Esquema de diferentes relatórios e fontes em um único dashboard (MICROSOFT, 2021)

3 MÉTODO

3.1 Descrição da Base de Dados

Partindo do arquivo “Tabela de Poços” (ANP, 2021), a triagem se iniciou analisando informações de 30.064 poços distribuídos em 37 bacias geológicas de 24 estados brasileiros. Cruzando o conteúdo desse material com a base de dados física adquirida no início de 2021, foram filtrados 21.307 poços existentes nela e que também estavam presentes e identificados na tabela. Desses, sete estavam duplicados ou não possuíam correspondência, totalizando 21.300 poços analisados através de nossa sistemática.

Os arquivos dessas dezenas de milhares de poços possuem formatos e extensões diferentes, a saber: .LIS, .DLIS, .PDF, .TXT, .TIFF e .LAS. Desse conjunto foram momentaneamente selecionados apenas poços que continham arquivos no formato .LIS e .DLIS. Os arquivos do tipo .LAS também se mostram de grande valia para trabalhos futuros, porém nesse primeiro momento possuem amostragem pouco significativa, aproximadamente 400 poços, em comparação aos milhares de poços com arquivos .LIS e .DLIS.

3.2 Sistemática Metodológica

O arquivo “Tabela de Poços” disponibilizado pela ANP, possui diversos conteúdos relativos aos poços *onshore* brasileiros, exemplificados na Tabela 1, para o poço 2-BTST-1-AM, da Bacia do Amazonas.

Tabela 1 – Dados do Poço 2-BTST-1-AM obtidos no arquivo Tabela de Poços (ANP, 2016)

Parâmetro	Valor
POCO	2-BTST-1-AM
CADASTRO	140200010100
POCO_OPERADOR	2BTST0001 AM
ESTADO	AM
BACIA	Solimões
BLOCO	Brasil
TERRA_MAR	T

POCO_POS_ANP	N
TIPO	Exploratório
CATEGORIA	Estratigráfico
RECLASSIFICACAO	SECO SEM INDÍCIOS
SITUACAO	ABANDONADO PERMANENTEMENTE
INICIO	21952
TERMINO	22097
CONCLUSAO	22097
TITULARIDADE	Público
LATITUDE_BASE_4C	-04:23:46,425
LONGITUDE_BASE_4C	-69:56:57,133
LATITUDE_BASE_DD	-4,3962291666
LONGITUDE_BASE_DD	-69,9492036111
DATUM_HORIZONTAL	SIRGAS2000
TIPO_DE_COORDENADA_DE_BASE	Definitiva
DIRECAO	Vertical
PROFUNDIDADE_SONDADOR_M	1482,3
REFERENCIA_DE_PROFUNDIDADE	MR
MESA_ROTATIVA	80
AGP	Existe
PC	Existe
PERFIS_DIGITAIS	Existe
GEOQUIMICA	Existe
SIG_SONDA	SC-34
NOM_SONDA	SONDA CONVENCIONAL 34
DHA_ATUALIZACAO	44143,2084

Fonte: ANP (2016)

Partindo da “Tabela de Poços” da ANP com a utilização de scripts em Python, foi desenvolvida a automação do processo de busca e catalogação de novos dados se utilizando de ferramentas conhecidas de mineração de dados quando necessário, a depender da evolução do reconhecimento dos tipos de dados necessários, como por exemplo REGEX (*Regular Expressions*) e JSON (*JavaScrip Object Notation*). Pretendeu-se assim indexar toda a abrangência de dados da “Tabela de Poços” com as curvas de perfil de poço extraídos dos arquivos DLIS e LIS.

Ainda existem, dentro do conteúdo disponibilizado pela ANP, diversos outros formatos de arquivos, como .PDF, .TIFF e .LAS, que não foram abordados nesse trabalho devido à baixa abrangência e maior complexidade para mineração dos dados.

3.2.1 Estruturação da base de dados não relacional em Python

Os arquivos do tipo LIS e DLIS podem se tornar difíceis de se trabalhar devido ao formato ter 30 anos de existência, com diferentes pacotes de software adicionando novos estruturas internas e tipos de objetos associados à esses arquivos. Esses arquivos contém grandes quantidades de metadados associados aos poços, assim como os próprios dados de poço que estão associados aos *Frames* e *Channels*. Os *Frames* podem representar diferentes passadas/corridas ou estágios de processamento (Bruto ou Interpretado) de dados, são tabelas que contém os dados de poço, onde cada coluna representa uma curva de perfilagem, e os dados são indexados por tempo ou profundidade. Cada curva (*logging)* de um *frame* é um canal (*Channel*), que pode ser unidimensional ou multi-dimensional.

Cada canal, ou curva de *logging*, associado aos frames de cada poço possui uma sigla específica (também conhecida como mnemônico), a qual não necessariamente padronizada, podendo variar de acordo com a empresa em que o perfil é adquirido, como por exemplo: Raios Gama (GR/RG/GRC), Densidade (Rhoz/Rhob/DEN), Fator Fotoelétrico (PE/PEF/PEF8). Foi objetivo deste trabalho identificar os mnemônicos mais relevantes (dado um universo de mais de 14 mil mnemônicos únicos somente nessa base) através de um sistema de triagem de variáveis e unidades de medida e posterior agrupamento (Tabela 2), visando facilitar as aquisições futuras de dados.

Tabela 2 - Exemplo de grupos de mnemônicos coletados

Grupo	**Mnemônico**
Caliper	RH, CALI, HCAL, CAL, CAL1,CAL2,CAL3,CAL4
Densidade	RHOZ, ZDEN, THOB, THOC, MD5

Afim de representar esses diferentes dados em um formato unificado que permita escalabilidade, flexibilidade e facilidade de armazenamento de dados com diferentes listas de variáveis (à exemplo das curvas de *logging* que variam de poço para poço, tanto em tipo quanto e quantidade) se escolheu o formato JSON (JavaScript Object Notation), que permite a construção rápida e intuitiva de *queries* e possui versatilidade de leitura com vários softwares e linguagens de programação.

Esse formato é descrito na forma de chave-valor, no qual os valores dos dados são acessados de acordo com a chave que os guarda.

Ao utilizar o Python e outras linguagens de programação como ferramentas de mineração de dados temos a possibilidade de criar dicionários e *queries* de fácil utilização e entendimento e assim permitir que o usuário faça suas consultas de maneiras dinâmica e interativa.

3.2.2 Captura de Informações com o uso de Regular Expression (*Regex*)

Utilizando *Regex (Regular Expressions)* foi possível buscar em arquivos .TXT, .DLIS e .LIS expressões regulares ou recorrentes como a bacia explorada em questão, a localização do poço em coordenadas geográficas ou até mesmo a profundidade da sonda e a direção do poço ou data de início e eventual fim de exploração e produção.

Tendo como *output* do código todas essas variáveis em formato de dicionário, torna-se de fácil análise e busca uma informação específica de cada um dos poços sem a necessidade de abrir cada uma das pastas ou arquivos, tornando eficiente e agradável a experiência do utilizador.

3.2.3 Limpeza de informações espúrias

A captura e processamento de dados precisou passar por alguns filtros e sistemáticas de limpeza e comunização de fontes. Apesar de encontrar 21.307 poços em nossa base de dados física, foram encontrados 8.371 poços com .DLIS e 7.168 poços com .LIS.

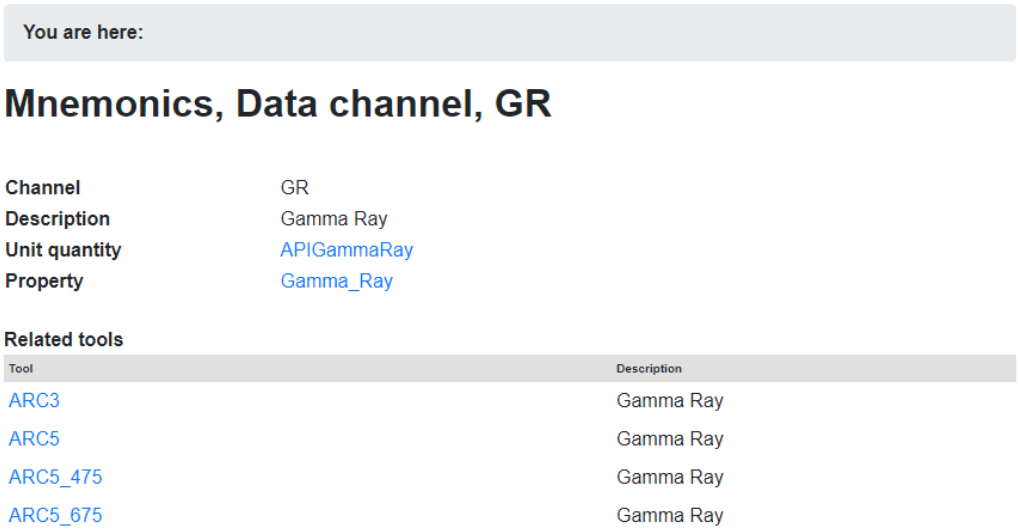
A partir dos dados capturados, foram removidos os poços duplicados, tais como os poços únicos que estavam sendo contabilizados duas vezes por divergência de caracteres. Algumas *strings* foram modificadas para se readequar ao padrão proposto de nomenclatura (vide tabela 3). Também foram descartados poços não identificados no ponto de partida.

Tabela 3 - Exemplos de poços com alterações para padronização

Base de Dados	Tabela ANP	Alteração proposta
1-FZB-131-CE	Ausente	Descarte
7-AG-0130-BA	7-AG-130-BA	Nomenclatura padronizada
7-AG-0140-BA	7-AG-140-BA	Nomenclatura padronizada
7-AG-0170-BA	7-AG-170-BA	Nomenclatura padronizada
7-AG-0180-BA	7-AG-180-BA	Nomenclatura padronizada
7-C-220-BA	7-C-220P-BA	Nomenclatura padronizada
7-ET-0470-RN	7-ET-470-RN	Nomenclatura padronizada
7-FZB-0480-CE	7-FZB-480-CE	Nomenclatura padronizada
7-FZB-580-CE	7-FZB-580D-CE	Nomenclatura padronizada
7-PIR-147-AL	7-PIR-147D-AL	Nomenclatura padronizada
7-RUC-57D-AM	Ausente	Descarte
7-SE-23A-RN	7-SE-23DA-RN	Nomenclatura padronizada
7-ET-0677-RN	7-ET-677-RN	Nomenclatura padronizada

3.2.4 Extração de mnemônicos

Os mnemônicos para cada perfil podem variar com a empresa responsável pela perfilagem. A extraçao dos mnemônicos, portanto, foi feita com o suporte de dicionários de empresas atuantes no segmento, como no caso do Dicionário de Curvas Mnemônicas da Schlumberger (SCHLUMBERGER, 2021) na Figura 5. Após o levantamento baseado na ocorrência de cada um deles, foi determinado um conjunto inicial para posterior agrupamento com base na frequência de aparição para usuários da indústria do petróleo e gás. Foi extraído do banco de metadados um arquivo .XLSX onde, após as primeiras colunas contendo a identificação do poço, *frame* e *channel* de cada leitura da triagem, tem-se um mnemônico por coluna e, como ocorrência de um metadado existente, a sinalização com um caractere singular (“1”).


You are here:

Mnemonics, Data channel, GR

Channel	GR
Description	Gamma Ray
Unit quantity	APIGammaRay
Property	Gamma_Ray

Related tools

Tool	Description
ARC3	Gamma Ray
ARC5	Gamma Ray
ARC5_475	Gamma Ray
ARC5_675	Gamma Ray

Figura 5 - Extração de tela da utilização do dicionário de curvas mnemônicos da Schlumberger Acesso em 20 de nov. de 2021.

Foram identificados 1.266 mnemônicos nos arquivos .LIS e 13.548 mnemônicos nos arquivos .DLIS, onde 746 destes eram comuns a ambas as integrações. Com isso, foram catalogados 14.068 mnemônicos únicos em todos os arquivos processados.

Tabela 4 - Número de ocorrências para os 30 mnemônicos mais frequentes da base

Mnemônico	Ocorrências
TDEP	52.089
TIME	38.234
GR	29.830
TENS	26.817
CALI	24.392
NPHI	21.493
DEPT	21.177
RHOB	19.742
DRHO	19.683
SP	19.646
BS	18.604
DUMM	14.705
CS	12.386
DT	10.344
ETIM	9.916
MSFL	8.890
AHVT	8.870
BHVT	8.866
IHV	7.524
ICV	7.246

BHV	7.238
AHV	7.234
GR	7.052
CILD	6.936
ILD	6.830
HCAL	6.812
HDRA	6.253
RHOZ	6.247
MINM	6.187
ITT	5.480

Em um universo com mais de 14 mil poços processados com arquivos .LIS e .DLIS e mais de 14 mil mnemônicos únicos, seria necessário uma planilha com cerca de 200 milhões de células para migrar todos os dados de plataforma.

Após o agrupamento de mnemônicos em 18 categorias de dados, foram contabilizados e considerados na extração de metadados 239 mnemônicos únicos. Esses, por sua vez, cobrem 1.184.282 aparições em nosso banco de dados, o que representa 22% do total. A intenção é que, ao longo do tempo, seja possível expandir e unificar esse dicionário de mnêmonicos e torná-lo cada vez mais abrangente.

3.2.5 Unificação da base de dados e visualização gráfica

Após a conclusão da base de dados dos perfis de poço presentes nos arquivos .DLIS e .LIS e a classificação de todos os mnemônicos relevantes, o projeto seguiu para um estudo de viabilidade e alternativas para análise exploratória e vizualização de dados, através de um *dashboard* com plotagens interativas. O manuseio dos dados permitiu obter uma ferramenta onde é possível ter um panorama geral dos dados catalogados para eventuais consultas. Para isso, parte do código direcionou os esforços para a geração de *outputs* do dicionário através do Microsoft Excel, esses permitindo o uso de um software comercial, no caso o Microsoft Power BI (software escolhido), ou scripts em Python, através de bibliotecas Open Source de vizualização, como Plotly, Matplotlib, Seaborn ou Bokeh e processamento de dados, como Pandas, Scikit-Learn e SciPy.

4 RESULTADOS

4.1 Base de dados não relacional em Python

Para confirmar a suposição de que a base de dados relacional (tabela 1) disponibilizada pela ANP continha os mesmos dados que a base de dados física (HD fornecido pela ANP) e também para verificar a viabilidade e assertividade do código Python desenvolvido, o mesmo foi executado em todos os arquivos de poços existentes na base física.

Os resultados mostraram que, partindo do universo de 30.064 poços da “Tabela de Poços” da ANP, sendo 23.295 deles *onshore*, 21.300 poços foram identificados pelo indexador e, destes, 15.412 tiveram arquivos do tipo .DLIS e/ou .LIS capturados pela ferramenta. Após o processamento dos dados, que tiveram alguns insucessos devido ao tamanho dos arquivos ou formato não regular dos arquivos, que necessitariam uma investigação aprofundada não abordada nesse trabalho, foram compilados com sucesso dados de 14.084 poços *onshore* brasileiros (Figura 6).

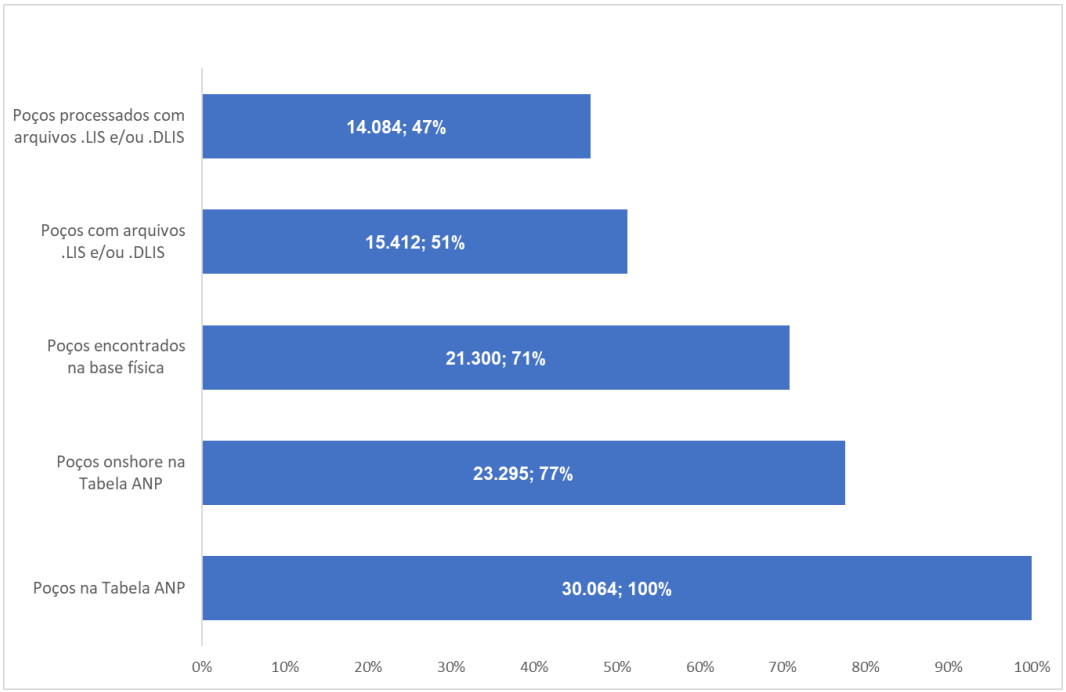


Figura 6 - Representação gráfica do número de poços processados em comparação ao total

O código visa disponibilizar as informações contidas nos arquivos do tipo .DLIS ou .LIS, ou seja, torna possível a análise de 14.084 poços com arquivos nesse formato. Foi notada uma capacidade de captura de dados de 91,4% dos poços onshore, onde mais de 60,5% deles tiveram seus perfis processados pelo JSON. Na Tabela 5 observa-se a quantidade de poços processados com arquivos em ambas as extensões por bacias:

Tabela 5 - Número de poços com dados processados por bacia

Bacia	Poços processados
Potiguar	5.852
Sergipe	3.832
Recôncavo	2.486
Espírito Santo	1.014
Alagoas	266
Solimões	205
Amazonas	144
Paraná	101
Tucano Sul	75
São Francisco	39
São Luís	18
Marajó	16
Tucano Central	9
Acre	8
Parecis - Alto Xingu	5
Rio do Peixe	5
Tucano Norte	3
Araripe	2
Tacutu	2
Campos	1
Jatobá	1
TOTAL	**14.084**

Durante a execução, foram encontrados 35 poços com nomes divergentes entre as duas bases. Para resolver essa questão, a descrição do poço foi alterada na base relacional da ANP para que coincidisse com a da base de dados física. O código retornou, também, 4 poços duplicados e 3 poços sem informações da base da ANP.

Ao capturar os dados dos arquivos .LIS, tem-se o total de 7.151 poços, porém existem 7.640 arquivos com essa extensão. Isso acontece por que há alguns poços que possuem mais de um arquivo .LIS que representam diferentes testes de poços, como é o caso do “'1-BRSA-276-RN”, que possui 16 versões diferentes. Ao todo, são 54 poços com mais de 1 arquivo. Além disso, refinou-se a busca para as pastas “Perfis

digitais” e “Perfil Convencional”, que são as localizações previstas pela padronização ANP, o que nos fornece um total de 6.673 poços assim dispostos na Tabela 6.

Tabela 6 - Número de poços com arquivos .LIS e sucesso no processamento por bacia

Bacia	Poços processados com .LIS	Sucesso no processamento
Potiguar	2.857	92,9%
Recôncavo	902	98,0%
Sergipe	2.563	98,9%
Espírito Santo	750	40,8%
Alagoas	6	66,7%
Solimões	12	58,3%
Amazonas	39	41,0%
Paraná	8	75,0%
São Francisco	13	53,8%
Rio do Peixe	1	0,0%
TOTAL	**7.151**	**89,7%**

A busca por arquivos do tipo .DLIS retornou um total de 20.063 arquivos de poços para apenas 8.370 poços. Da mesma forma que acontece com arquivos .LIS, é possível que um único poço tenha dezenas de arquivos distintos com extensão .DLIS. Para ilustrar esse ponto, a Tabela 7 mostra os 10 poços com a maior quantidade de arquivos .DLIS:

Tabela 7 - Dez poços com o maior número de arquivos .DLIS

Descrição do poço	Quantidade de arquivos .DLIS
1-ELPS-4-PR	91
1-COST-1P-PR	51
1-GOP-1A-BA	36
7-SBO-1-RN	35
7-SDM-1-RN	29
7-SBO-2-RN	26
1-REV-1-BA	23
1-IMET-8-MG	20
2-ANP-6-MT	19
8-SDM-3-RN	17

No caso dos arquivos .DLIS, ao refinar a busca para pastas específicas de maneira idêntica ao que foi realizado nos .LIS, o número de arquivos cai para 12.005, distribuídos nos mesmos 8.370 poços pelas bacias da seguinte forma (Tabela 8):

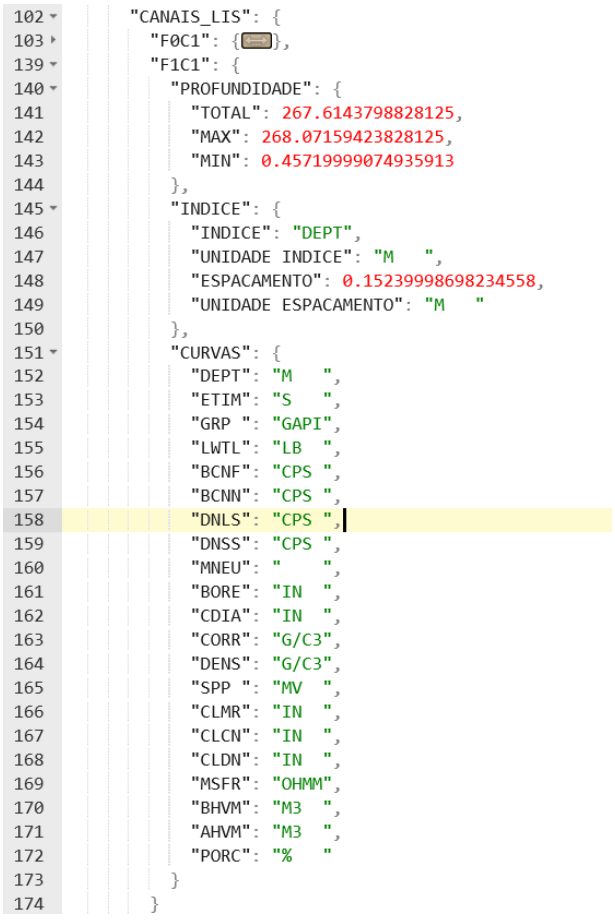
Tabela 8 - Número de poços com arquivo .DLIS e sucesso no processamento por bacia

Bacia	Poços processados com .DLIS	Sucesso no processamento
Potiguar	3.291	95,7%
Recôncavo	1.644	97,0%
Sergipe	1.346	95,1%
Espírito Santo	855	80,7%
Alagoas	387	68,7%
Solimões	275	73,8%
Amazonas	142	98,6%
Parnaíba	143	0,0%
Tucano Sul	82	91,5%
Paraná	98	96,9%
São Francisco	40	97,5%
São Luís	18	100,0%
Marajó	16	100,0%
Tucano Central	9	100,0%
Acre	8	100,0%
Parecis-Alto Xingu	5	100,0%
Rio do Peixe	5	100,0%
Tacatu	2	100,0%
Jatobá	1	100,0%
Araripe	2	100,0%
Campos	1	0,0%
TOTAL	**8.370**	**90,8%**

As informações captadas da tabela de poços da ANP, em conjunto com os mnemônicos minerados nos arquivos de extensão .DLIS ou .LIS podem ser observados no formato JSON não relacional na Figura 7. Os arquivos .LIS e .DLIS, assim como cada corrida presente internamente nesses arquivos, foram nomeados no formato FXXCYY, onde os primeiros dígitos se referem ao número do arquivo do poço e os últimos a qual corrida de captação de dados interna ao arquivo àqueles dados se referem. Na Figura 8 temos um exemplo da leitura dos canais no dicionário onde é possível identificar, para cada parâmetro, o mnemônico e sua respectiva unidade de medida, que foram convertidas para permitir o agrupamento de mnemônicos por conteúdo.


```
{
  "Potiguar": {
    "Poti_1-BRSA-29-RN_Pionei_T": {
      "CADASTRO": 72100019112,
      "OPERADOR": "Potiguar E&P S.A.",
      "POCO_OPERADOR": "1NVRG1RN",
      "ESTADO": "RN",
      "BACIA": "Potiguar",
      "BLOCO": "BPOT-4",
      "TERRA_MAR": "T",
      "POCO_POS_ANP": "S",
      "TIPO": 1,
      "CATEGORIA": "Pioneiro",
      "RECLASSIFICACAO": "PORTADOR DE PETRÓLEO",
      "SITUACAO": "ABANDONADO PERMANENTEMENTE",
      "INICIO": "2000-11-20 00:00:00",
      "TERMINO": "2000-12-19 00:00:00",
      "CONCLUSAO": "2001-04-08 00:00:00",
      "TITULARIDADE": "Público",
      "LATITUDE_BASE_4C": "-05:32:48,432",
      "LONGITUDE_BASE_4C": "-37:38:11,328",
      "LATITUDE_BASE_DD": "-5,5467866666",
      "LONGITUDE_BASE_DD": "-37,63648",
      "DATUM_HORIZONTAL": "SIRGAS2000",
      "TIPO_DE_COORDENADA_DE_BASE": "Definitiva",
      "DIRECAO": "Vertical",
      "PROFUNDIDADE_SONDADOR_M": 2326,
      "REFERENCIA_DE_PROFUNDIDADE": "MR",
      "MESA_ROTATIVA": 76,
      "COTA_ALTIMETRICA_M": 67.3,
      "LAMINA_D_AGUA_M": 0,
      "DATUM_VERTICAL": "NM",
      "CDPE": "Existe",
      "AGP": "Existe",
      "PC": "Existe",
      "PERFIS_DIGITAIS": "Existe",
      "SIG_SONDA": "SC-106",
      "NOM_SONDA": "SONDA CONVENCIONAL 106",
      "DHA_ATUALIZACAO": "2021-08-08 05:00:07",
```


Figura 7 - Exemplo de poço no formato não relacional JSON




```
      "CANAIS_LIS": {
        "F0C1": {...},
        "F1C1": {
          "PROFUNDIDADE": {
            "TOTAL": 267.6143798828125,
            "MAX": 268.07159423828125,
            "MIN": 0.45719999074935913
          },
          "INDICE": {
            "INDICE": "DEPT",
            "UNIDADE INDICE": "M   ",
            "ESPACAMENTO": 0.15239998698234558,
            "UNIDADE ESPACAMENTO": "M   "
          },
          "CURVAS": {
            "DEPT": "M   ",
            "ETIM": "S   ",
            "GRP ": "GAPI",
            "LWTL": "LB  ",
            "BCNF": "CPS ",
            "BCNN": "CPS ",
            "DNLS": "CPS ",
            "DNSS": "CPS ",
            "MNEU": "    ",
            "BORE": "IN  ",
            "CDIA": "IN  ",
            "CORR": "G/C3",
            "DENS": "G/C3",
            "SPP ": "MV  ",
            "CLMR": "IN  ",
            "CLCN": "IN  ",
            "CLDN": "IN  ",
            "MSFR": "OHMM",
            "BHVM": "M3  ",
            "AHVM": "M3  ",
            "PORC": "%   "
          }
        }
```


Figura 8 - Formato dos canais após mineração dos dados

4.2 Análise, categorização e agrupamento de mnemônicos da base não-relacional

Foram identificados 18 grupos de mnemônicos mais comuns ao profissional que atua com o ramo de perfilagem. Os grupos foram desenvolvidos e verificados a partir das unidades de medida (já convertidas, quando necessário). Os grupos de mnemônicos estão reportados na Tabela 9.

Tabela 9 - Mnemônicos agrupados de acordo com demanda e frequência

Grupo de Mnemônicos
Anisotropia Acústica
Anisotropia Resistiva
Bit Size (Diâmetro da Broca)
Caliper (Diâmetro do Poço)
Densidade
Fator Fotoelétrico
Gamma Ray
Gamma Ray Espectral
Litogeoquímico
Microrresistividade
Porosidade
Potencial Espontâneo
Profundidade
Resistividade
Resmag (Resistência Magnética)
Sônico
Velocidade de Cabo
Volume do Poço

De posse dos 18 grupos de mnemônicos, foi possível gerar um único arquivo – “Lista de canais com mnemônicos agrupados” - de proporções razoáveis (36,8 MB) e que permitisse ao utilizador não depender de computadores de alta performance para trabalhar com os metadados.

Esse arquivo, de formato .XLSX foi extraído através da programação em Python, por meio da consulta e compilação desses dados da base JSON e posteriormente colocado no formato tabular de forma a permitir fácil acesso a esses dados para todos os usuários. Tais utilizadores, mesmo que sem acesso a ferramentas como o MatLab ou Power BI, podem analisar os metadados dos poços processados em suas atividades científicas e tecnológicas, sejam estas acadêmicas ou industriais.

O resultado da indexação de metadados por poços foi distribuído em uma planilha com quase 66 mil entradas, separados linha a linha com base em uma ID_interna, utilizada para identificar o poço na base unificada e padronizada. As variáveis associadas a cada ID dizem respeito à categoria do poço, sua respectiva bacia e eventual campo, assim como o número de *frames* e corridas para posterior leitura dos dados e profundidade de cada perfil adquirido. Somam-se a essas informações 18 colunas, uma para cada categoria de mnemônicos. Os parâmetros foram, portanto, reunidos de acordo com o exemplo reduzido da Figura 9.

#	ID_INTERNA	POCO	CAT	TIPO	BACIA	CAMPO	FRAME	CORRIDA	PMIN	PMAX	PTOTAL	Micro-Resistividade	Sônico	Anisotropia Resistiva	Fator Fotoeletrico	Litogeoquímico	Velocidade Cabo	Profundidade	Resistividade	Gamma Ray
220	Alag_7-PIR-245D-AL_Desenv_T	7-PIR-245D-AL	7	Desenvolvimen	Alagoas	PILAR	F1	C6	2.949,91	3.054,10	104,19	1			1			1		1
221	Alag_7-ANB-13D-AL_Desenv_T	7-ANB-13D-AL	7	Desenvolvimen	Alagoas	ANAMBÉ	F0	C0										1		1
222	Alag_7-ANB-13D-AL_Desenv_T	7-ANB-13D-AL	7	Desenvolvimen	Alagoas	ANAMBÉ	F1	C0										1		1
223	Alag_3-PIA-23-AL_Extens_T	3-PIA-23-AL	3	Extensão	Alagoas	PIAÇABUÇU	F0	C0	0,00	3,00	3,00									
224	Alag_3-PIA-23-AL_Extens_T	3-PIA-23-AL	3	Extensão	Alagoas	PIAÇABUÇU	F1	C0	0,00	3,00	3,00									
225	Alag_3-PIA-23-AL_Extens_T	3-PIA-23-AL	3	Extensão	Alagoas	PIAÇABUÇU	F2	C0	0,00	3,00	3,00									
226	Alag_3-PIA-23-AL_Extens_T	3-PIA-23-AL	3	Extensão	Alagoas	PIAÇABUÇU	F3	C0	0,00	3,00	3,00									
227	Alag_3-PIA-23-AL_Extens_T	3-PIA-23-AL	3	Extensão	Alagoas	PIAÇABUÇU	F4	C0	328,50	6.059,00	5.730,50						1			
228	Alag_3-PIA-23-AL_Extens_T	3-PIA-23-AL	3	Extensão	Alagoas	PIAÇABUÇU	F5	C0	295,50	6.059,00	5.763,50		1				1		1	1
229	Alag_3-PIA-23-AL_Extens_T	3-PIA-23-AL	3	Extensão	Alagoas	PIAÇABUÇU	F6	C0	2.030,50	6.056,00	4.025,50									1
230	Alag_3-PIA-23-AL_Extens_T	3-PIA-23-AL	3	Extensão	Alagoas	PIAÇABUÇU	F7	C0	2.060,50	6.040,00	3.979,50						1			
231	Alag_3-SCE-4-AL_Extens_T	3-SCE-4-AL	3	Extensão	Alagoas	SUL DE CORURIPE	F0	C0	0,00	3,00	3,00						1			
232	Alag_3-SCE-4-AL_Extens_T	3-SCE-4-AL	3	Extensão	Alagoas	SUL DE CORURIPE	F1	C0	0,00	3,00	3,00						1			
233	Alag_3-SCE-4-AL_Extens_T	3-SCE-4-AL	3	Extensão	Alagoas	SUL DE CORURIPE	F2	C0	0,00	3,00	3,00						1			
234	Alag_3-SCE-4-AL_Extens_T	3-SCE-4-AL	3	Extensão	Alagoas	SUL DE CORURIPE	F3	C0	1.245,00	1.771,80	526,80						1			1
235	Alag_3-SCE-4-AL_Extens_T	3-SCE-4-AL	3	Extensão	Alagoas	SUL DE CORURIPE	F4	C0	0,00	3,00	3,00									
236	Alag_3-SCE-4-AL_Extens_T	3-SCE-4-AL	3	Extensão	Alagoas	SUL DE CORURIPE	F5	C0	20,00	747,80	727,80		1						1	1
237	Alag_3-SCE-4-AL_Extens_T	3-SCE-4-AL	3	Extensão	Alagoas	SUL DE CORURIPE	F6	C0	744,00	1.770,80	1.026,80	1	1						1	1
238	Alag_3-SCE-4-AL_Extens_T	3-SCE-4-AL	3	Extensão	Alagoas	SUL DE CORURIPE	F7	C0	748,60	1.775,80	1.027,20								1	
239	Alag_3-RFA-4-AL_Extens_T	3-RFA-4-AL	3	Extensão	Alagoas	0	F0	C0	0,00	3,00	3,00									

Figura 9 - Captura de tela para ilustração da lista de mnemônicos agrupados

4.3 Sistema de triagem de metadados via *dashboard*

Ao converter os dados consolidados na “Lista de canais com mnemônicos agrupados” para Excel, foi desenvolvido um primeiro exemplo de *dashboard* (Figura 10), onde o usuário final tem a possibilidade de pesquisar por certas áreas de interesse através dos mnemônicos presentes em cada curva. Esse cenário possibilita que o operador do material possa restringir suas buscas por um conceito, variável ou área específica de interesse visando localizar um caso de estudo em particular.

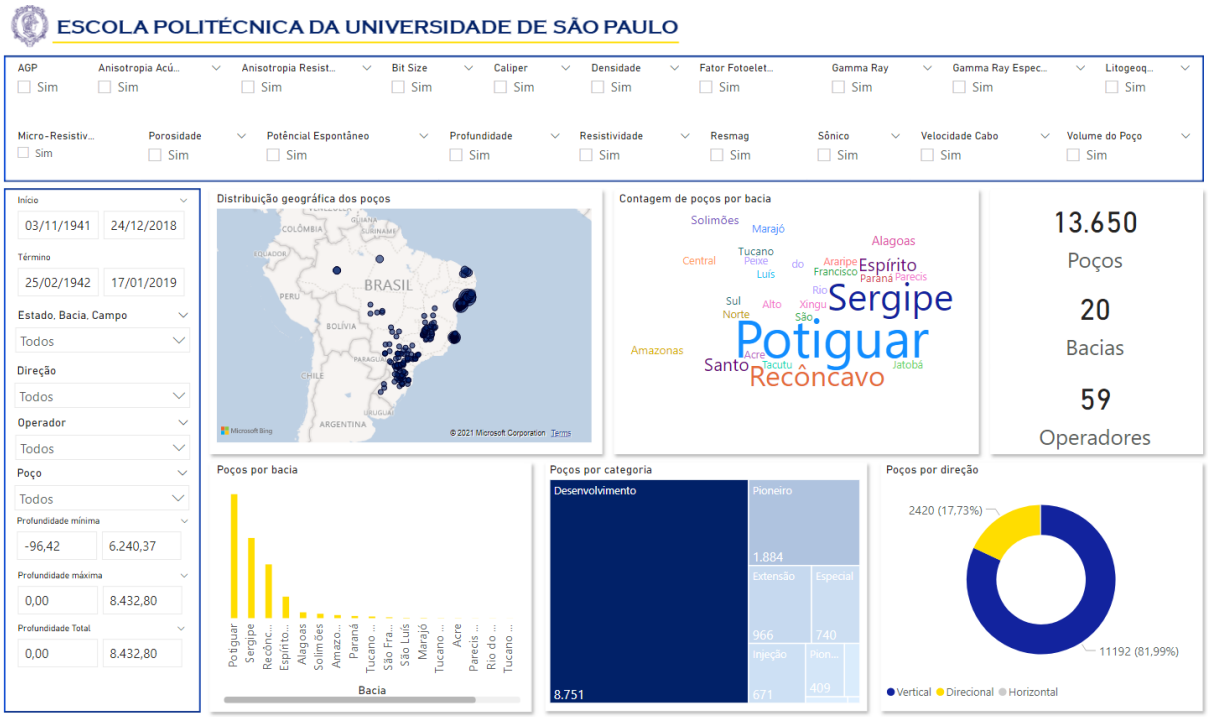


Figura 10 - Cenário de dashboard visando a busca por mnemônicos de interesse

Também é possível abordar um cenário (Figura 11) onde o usuário tenha a opção de pesquisar por um nome de poço específico, categoria de poço ou até mesmo bacia ou operador. Neste caso, é possível restringir inúmeras variáveis e verificar diversas condições específicas de análises de perfilagem de poços das bacias continentais brasileiras ou até mesmo de outras bases de dados originais e inéditas.

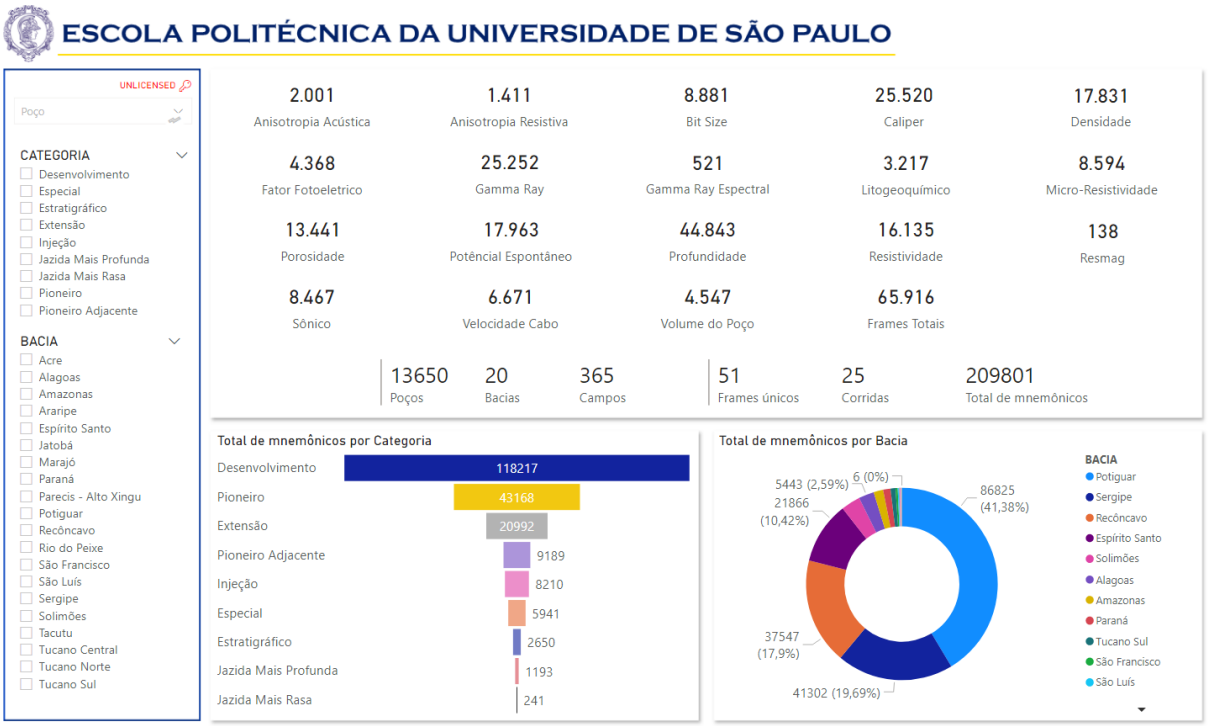


Figura 11 - Cenário de dashboard visando a busca por especificidade e gama de mnemônicos

5 CONCLUSÃO

O projeto foi composto por três pilares básicos, a saber:

i. Código indexador do acervo de dados para transformá-los em um banco não-relacional, no formato .JSON estará disponível no acervo da disciplina TCC 2 da Escola Politécnica.
ii. Análise, categorização e agrupamento de mnemônicos, que visou garantir uma factibilidade ao trabalho fazendo todo o processo de triagem dos mnemônicos de interesse nesse primeiro momento de pesquisa;
iii. Ferramenta de visualização em *dashboards* e dicionários, facilitando a futura interpretação dos dados e metadados antes compilados em arquivos dos mais diversos formatos em pastas e subpastas.

A reunião destes três pilares possibilitou a otimização do processo de busca por poços e suas respectivas assembleias de perfis. O trabalho desenvolvido favorece buscas simples e eficientes, o que propicia a redução de hora humana trabalhada em processos simples de localização de perfis associados aos poços, sejam estes exploratórios ou produtores.

5.1 Contribuições do trabalho

Este trabalho dará suporte a futuras pesquisas e consultas à base de dados de poços *onshore* brasileiros, bem como introduz um conceito simples que pode ser expandido também aos poços das bacias *offshore*. O conceito permite a rápida e precisa triagem e coleta de metadados dentro de inúmeras bases de dados de arquivos de poços nos mais diversos formatos (.TXT, .PDF, .DLIS, .LIS, .LAS), e permite transportar essas informações para um ambiente de busca e visualização gráfica, como por exemplo através da geração de novos *dashboards* e apresentações.

5.2 Trabalhos futuros

Ainda diante do vasto material disponibilizado, é possível explorar mais a fundo o repertório de siglas não-padronizadas (mnemônicos) presentes nos arquivos .DLIS. Como cada operadora possui seu próprio dicionário de siglas, é possível explorar ferramentas de análise semântica automatizadas para encontrar agrupamentos de parâmetros similares e mais complexos do que os que foram abordados nesse trabalho.

No perímetro de visualização de dados é simples e factível agregar conteúdos de plotagem interativa, clusterização e ambientes dashboard das mais diversas formas para auxiliar na interpretação das informações contidas em uma base original já processada, criando um panorama geral dos dados catalogados através dessas ferramentas em eventuais consultas de terceiros.

REFERÊNCIAS BIBLIOGRÁFICAS

AGÊNCIA NACIONAL DO PETRÓLEO, GÁS NATURAL E BIOCOMBUSTÍVEIS - SUPERINTENDÊNCIA DE DADOS TÉCNICOS – SDT. **Nota Técnica nº 073/2016/SDT.** 2016.

ANGWIN, M.; NELSM, J. L.; SYRETT, B. C. 1996. **Technical Database Design**. The NACE International Annual Conference and Exposition. 1996.

ANP. **Acervo de Dados**. Disponível em: <http://www.anp.gov.br/exploracao-e-producao-de-oleo-e-gas/dados-tecnicos/acervo-de-dados>. Acesso em 22 de jul. de 2021.

CODD, E. F. A Relational Model of Data for Large Shared Data Banks. **Communications of the ACM**, v. 13, n. 6, p. 377–387, 1 jun. 1970.

COX, N. Exploratory Data Mining and Data Cleaning. **Journal of Statistical Software**, v. 11, 27 out. 2004.

Data B Inteligência. **O que é Power BI e por que usar em seu negócio?** Disponível em: <https://databinteligencia.com.br/o-que-e-power-bi-e-por-que-usar-em-seu-negocio/>. Acesso em 21 de nov. de 2021

DE ANDRADE, P. A. **Assinatura de Variáveis em Perfis de Poços na determinação de zonas de interesse para óleo e gás com base em Mapas Auto-Oganizáveis (SOM)**. 2019.

ENDEL, F.; PIRINGER, H. **Data Wrangling: Making data useful again**. IFAC-PapersOnLine. Elsevier, 1 fev. 2015

FRACZEK, K.; PLECHAWSKA-WOJCIK, M. **Comparative Analysis of Relational and Non-relational Databases in the Context of Performance in Web Applications**. 153-164. 10.1007/978-3-319-58274-0_13. 2017.

Informática UFSC. **Conceito Básicos da Teoria de Grafos**. Disponível em:< https://www.inf.ufsc.br/grafos/definicoes/definicao.html>. Acesso em 21 de nov. de 2021.

JATIN; BATRA, S. MONGODB Versus SQL: A Case Study on Electricity Data. p. 297–308. 2016

KANDEL, S. et al. Research directions in data wrangling: Visualizations and transformations for usable and credible data. **Information Visualization**, v. 10, n. 4, p. 271–288, 2 out. 2011.

Microsoft Docs. **Noções básicas de dashboard.** Disponível em: <https://docs.microsoft.com/pt-br/power-bi/create-reports/service-dashboards>. Acesso em 19 de nov. de 2021.

PHIRI, H.; KUNDA, D. A Comparative Study of NoSQL and Relational Database. **Zambia ICT Journal**, v. 1, p. 1–4, 27 dez. 2017.

Schlumberger. **Curve Mnemonic Dictionary.** Disponível em: <https://www.apps.slb.com/cmd/index.aspx> Acesso em 19 de nov. de 2021.

WICKHAM, H. Tidy data. **Journal of Statistical Software**, v. 59, n. 10, p. 1–23, 1 set. 2014.

Universidade de São Paulo

Engenharia de Petróleo – Escola Politécnica

Número: 8585561USP **Data: 25/11/2021**

ESTRUTURAÇÃO DE BANCO DE DADOS NÃO-RELACIONAL PARA AS INFORMAÇÕES DE POÇOS DE PETRÓLEO DAS BACIAS SEDIMENTARES CONTINENTAIS BRASILEIRAS

Gustavo Lira Girardi de Góes

Orientador: Prof. Dr. Cleyton de Carvalho Carneiro

Co-Orientador: Rodrigo César de Teixeira Gouvêa

Artigo Sumário referente à disciplina PMI3349 – Trabalho de Conclusão de Curso II
Este artigo foi preparado como requisito para completar o curso de Engenharia de Petróleo na Escola Politécnica da USP. Template versão 2021v01.

Resumo

O presente artigo apresenta, por meio de ferramentas computacionais, uma análise exploratória inicial dos dados e o desenvolvimento de um sistema indexador para triagem das informações de poços das bacias *onshore* brasileiras, permitindo a observação e análise prévia dos dados do conteúdo disponibilizado, e também dos dados de perfuração com os perfis disponíveis para mais de 21 mil poços por meio de uma base de dados não-relacional. Com mais de 14 mil poços processados com arquivos do tipo LIS e DLIS, é possível utilizá-los no processo de localização e catalogação de informações específicas em trabalhos que se utilizarão da base de dados.

Abstract

This work aims, through computational tools, to carry out the initial exploratory analysis of the data and develop an indexing system for screening the information from each well, allowing the observation and prior analysis of the data available, as well as the drilling data with the logs available for more than 21,000 wells through a non-relational database. With more than 14,000 wells processed with LIS and DLIS files, it is possible to use them for locating and cataloging specific information in works that will use the database.

1. Introdução

No início de 2021, a Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP) nos disponibilizou em mídia física diversos dados de exploração e produção relativos às bacias onshore brasileiras. Dentro desse conteúdo estão dados de poços que compõe um acervo relevante, com metadados diversos dotados de uma variedade de poços adquiridos por diversas companhias.

O conteúdo da base de dados disponibilizada engloba 21.307 poços distribuídos em 23 bacias geológicas de 22 estados brasileiros. Esses dados compactados resultam em mais de 1,3 TB de informação não integrada.

1.1. Objetivos

Um dos pilares desse projeto de mineração dos dados cedidos pela ANP são os arquivos com extensão DLIS (Digital Log Interchange Standard) e LIS (Log Interchange Standard). Os arquivos .DLIS e .LIS proporcionam um compilado de dados de poços extraídos a partir das

ferramentas de perfilagem como porosidade, raios gama, resistividade ressonância e diâmetro do poço.

1.2. Justificativa

A complexidade associada a compreensão das informações de poços é o objeto central deste projeto. Apesar de cada pasta de poço da base de dados disponibilizada originalmente conter os respectivos arquivos de extensão .DLIS, não é possível acessar ao conteúdo dos canais de perfilagem associado a cada poço sem a buscar item a item, pasta por pasta. Entende-se, portanto, a importância da compilação desses metadados em uma base indexada separada, que possa ser consultada de forma rápida e intuitiva para a aquisição e compilação dessas informações.

2. Metodologia

Partindo do arquivo "Tabela de Poços" (ANP, 2021), a triagem se iniciou analisando informações de 30.064 poços distribuídos em 37 bacias geológicas de 24 estados brasileiros. Cruzando o conteúdo desse material com a base de dados física adquirida no início de 2021, foram filtrados 21.307 poços existentes nela e que também estavam presentes e identificados na tabela. Desses, sete estavam duplicados ou não possuíam correspondência, totalizando 21.300 poços analisados através de nossa sistemática.

Os arquivos dessas dezenas de milhares de poços possuem formatos e extensões diferentes, a saber: .LIS, .DLIS, .PDF, .TXT, .TIFF e .LAS. Desse conjunto foram momentaneamente selecionados apenas poços que continham arquivos no formato .LIS e .DLIS.

2.1. Sistemática Metodológica

2.1.1. Estruturação da base de dados não relacional em Python

Cada canal, ou curva de logging, associado aos frames de cada poço possui uma sigla específica (também conhecida como mnemônico), a qual não necessariamente padronizada, podendo variar de acordo com a empresa em que o perfil é adquirido, como por exemplo: Gamma Ray (GR/RG/GRC), Densidade (Rhoz/Rhob/DEN), Fator Fotoelétrico (PE/PEF/PEF8). Foram identificados os mnemônicos mais relevantes (dado um universo de mais de 14 mil mnemônicos únicos somente nessa base) através de um sistema de triagem de variáveis e unidades de medida e posterior agrupamento (Tabela 1), visando facilitar as aquisições futuras de dados.

Tabela 10 - Exemplo de grupos de mnemônicos coletados

Grupo	Mnemônico

Caliper	RH, CALI, HCAL, CAL, CAL1,CAL2,CAL3,CAL4
Densidade	RHOZ, ZDEN, THOB, THOC, MD5

Afim de representar esses diferentes dados em um formato unificado que permita escalabilidade, flexibilidade e facilidade de armazenamento de dados com diferentes listas de variáveis (à exemplo das curvas de logging que variam de poço para poço, tanto em tipo quanto e quantidade) se escolheu o formato JSON (JavaScript Object Notation), que permite a construção rápida e intuitiva de queries e possui versatilidade de leitura com vários softwares e linguagens de programação. Esse formato é descrito na forma de chave-valor, no qual os valores dos dados são acessados de acordo com a chave que os guarda.

2.1.2. Captura de Informações com o uso de Regular Expression (Regex)

Utilizando Regex (Regular Expressions) foi possível buscar em arquivos .TXT, .DLIS e .LIS expressões regulares ou recorrentes como a bacia explorada em questão, a localização do poço em coordenadas geográficas ou até mesmo a profundidade da sonda e a direção do poço ou data de início e eventual fim de exploração e produção.

2.1.3. Limpeza de informações espúrias

A captura e processamento de dados precisou passar por alguns filtros e sistemáticas de limpeza e comunização de fontes. Apesar de encontrar 21.307 poços em nossa base de dados física, foram encontrados 8.371 poços com .DLIS e 7.168 poços com .LIS.

2.1.4. Extração de mnemônicos

Os mnemônicos para cada perfil podem variar com a empresa perfuradora. A extraçao dos mnemônicos, portanto, foi feita com o suporte de dicionários de empresas atuantes no segmento, como no caso do Dicionário de Curvas Mnemônicas da Schlumberger (SCHLUMBERGER, 2021). Após o levantamento baseado na ocorrência de cada um deles, foi determinado um conjunto inicial para posterior agrupamento com base na frequência de aparição para usuários da indústria do

petróleo e gás. Foi extraído do banco de metadados um arquivo .XLSX onde, após as primeiras colunas contendo a identificação do poço, frame e canal de cada leitura da triagem, tem-se um mnemônico por coluna e, como ocorrência de um metadado existente, a sinalização com um caractere singular (“1”).

Foram identificados 1.266 mnemônicos nos arquivos .LIS e 13.548 mnemônicos nos arquivos .DLIS, onde 746 destes eram comuns a ambas as integrações. Com isso, foram catalogados 14.068 mnemônicos únicos em todos os arquivos processados.

Após o agrupamento de mnemônicos em 18 categorias de dados, foram contabilizados e considerados na extração de metadados 239 mnemônicos únicos. Esses, por sua vez, cobrem 1.184.282 aparições em nosso banco de dados, o que representa 22% do total. A intenção é que, ao longo do tempo, seja possível expandir e unificar esse dicionário mnêmico e torná-lo cada vez mais abrangente.

2.1.5. Unificação da base de dados e visualização gráfica

Após a conclusão da base de dados dos perfis de poço presentes nos arquivos .DLIS e .LIS e a classificação de todos os mnemônicos relevantes, o projeto seguiu para um estudo de viabilidade e alternativas para análise exploratória e vizualização de dados, através de um dashboard com plotagens interativas. O manuseio dos dados permitiu obter uma ferramenta onde é possível ter um panorama geral dos dados catalogados para eventuais consultas.Para isso, parte do código direcionou os esforços para a geração de outputs do dicionário através do Microsoft Excel, esses permitindo o uso de um software comercial, no caso o Microsoft Power BI (software escolhido), ou scripts em Python, através de bibliotecas Open Source de vizualização, como Plotly, Matplotlib, Seaborn ou Bokeh e processamento de dados, como Pandas, Scikit-Learn e SciPy.

3. Resultados

3.1. Base de dados não-relacional em Python

Para confirmar a suposição de que a base de dados relacional disponibilizada pela ANP continha os mesmos dados que a base de dados física (HD fornecido pela ANP) e também para verificar a viabilidade e assertividade do código Python desenvolvido, o mesmo foi executado em todos os arquivos de poços existentes na base física.

Os resultados mostraram que, partindo do universo de 30.064 poços da “Tabela de Poços” da ANP, sendo 23.295 deles onshore, 21.300 poços foram identificados pelo indexador e,

destes, 15.412 tiveram arquivos do tipo .DLIS e/ou .LIS capturados pela ferramenta. Após o processamento dos dados, que tiveram alguns insucessos devido ao tamanho dos arquivos ou formato não regular dos arquivos, que necessitariam uma investigação aprofundada não abordada nesse trabalho, foram compilados com sucesso dados de 14.084 poços onshore brasileiros (Figura 1).

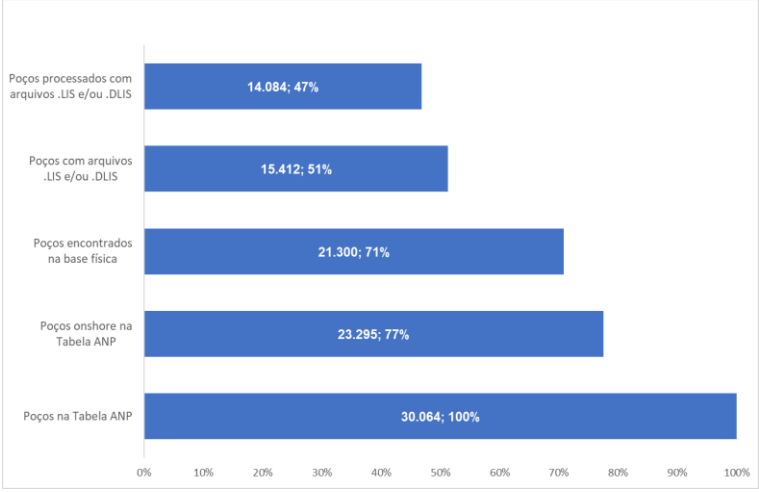


Figura 1 - Representação gráfica do número de poços processados em comparação ao total

Ao capturar os dados dos arquivos .LIS, tem-se o total de 7.151 poços, porém existem 7.640 arquivos com essa extensão. Isso acontece por que há alguns poços que possuem mais de um arquivo .LIS que representam diferentes testes de poços, como é o caso do "'1-BRSA-276-RN", que possui 16 versões diferentes. Ao todo, são 54 poços com mais de 1 arquivo. Além disso, refinou-se a busca para as pastas "Perfis digitais" e "Perfil Convencional", que são as localizações previstas pela padronização ANP, o que nos fornece um total de 6.673 poços assim dispostos na Tabela 2.

Tabela 11 - Número de poços com arquivos .LIS e sucesso no processamento por bacia

Bacia	Poços processados com .LIS	Sucesso no processamento
Potiguar	2.857	92,9%
Recôncavo	902	98,0%
Sergipe	2.563	98,9%
Espírito Santo	750	40,8%
Alagoas	6	66,7%
Solimões	12	58,3%
Amazonas	39	41,0%
Paraná	8	75,0%
São Francisco	13	53,8%
Rio do Peixe	1	0,0%
TOTAL	**7.151**	**89,7%**

No caso dos arquivos .DLIS, ao refinar a busca para pastas específicas de maneira idêntica ao que foi realizado nos .LIS, o número de arquivos cai para 12.005, distribuídos nos mesmos 8.370 poços pelas bacias da seguinte forma (Tabela 3):

Tabela 12 - Número de poços com arquivos .DLIS e sucesso no processamento por bacia

Bacia	Poços processados com .DLIS	Sucesso no processamento
Potiguar	3.291	95,7%
Recôncavo	1.644	97,0%
Sergipe	1.346	95,1%
Espírito Santo	855	80,7%
Alagoas	387	68,7%
Solimões	275	73,8%
Amazonas	142	98,6%
Parnaíba	143	0,0%
Tucano Sul	82	91,5%
Paraná	98	96,9%
São Francisco	40	97,5%
São Luís	18	100,0%
Marajó	16	100,0%
Tucano Central	9	100,0%
Acre	8	100,0%
Parecis-Alto Xingu	5	100,0%
Rio do Peixe	5	100,0%
Tacatu	2	100,0%
Jatobá	1	100,0%
Araripe	2	100,0%
Campos	1	0,0%
TOTAL	**8.370**	**90,8%**

3.2. Análise, categorização e agrupamento de mnemônicos da base não-relacional

Foram identificados 18 grupos de mnemônicos mais comuns ao profissional que atua com o ramo de perfilagem. Os grupos foram desenvolvidos e verificados a partir das unidades de medida (já convertidas, quando necessário). Os grupos de mnemônicos estão reportados na Tabela 4.

Tabela 13 – Exemplo de grupos de mnemônicos de acordo com demanda e frequência

Grupo de Mnemônicos
Anisotropia Acústica
Bit Size (Diâmetro da Broca)
Caliper (Diâmetro do Poço)
Densidade
Fator Fotoelétrico
Gamma Ray
Litogeoquímico
Microrresistividade
Porosidade
Potencial Espontâneo
Profundidade
Resistividade
Resmag
Sônico
Velocidade de Cabo

De posse dos 18 grupos de mnemônicos, foi possível gerar um único arquivo – “Lista de canais com mnemônicos agrupados” - de proporções razoáveis (36,8 MB) e que permitisse ao utilizador não depender de computadores de alta performance para trabalhar com os metadados.

Esse arquivo, de formato .XLSX foi extraído através da programação em Python, por meio da consulta e compilação desses dados da base JSON e posteriormente colocado no formato tabular de forma a permitir fácil acesso a esses dados para todos os usuários.

O resultado da indexação de metadados por poços foi distribuído em uma planilha com quase 66 mil entradas, separados linha a linha com base em uma ID_interna, utilizada para identificar o poço na base unificada e padronizada. As variáveis associadas a cada ID dizem respeito à categoria do poço, sua respectiva bacia e eventual campo, assim como o número de *frames* e corridas para posterior leitura dos dados e profundidade de cada perfil adquirido. Somam-se a essas informações 18 colunas, uma para cada categoria de mnemônicos.

3.3. Sistema de triagem de metadados via *dashboard*

Ao converter os dados consolidados na “Lista de canais com mnemônicos agrupados” para Excel, foi desenvolvido um primeiro exemplo de *dashboard* (Figura 2), onde o usuário final tem a possibilidade de pesquisar por certas áreas de interesse através dos mnemônicos presentes em cada curva. Esse cenário possibilita que o operador do material possa restringir suas buscas por um conceito, variável ou área específica de interesse visando localizar um caso de estudo em particular.

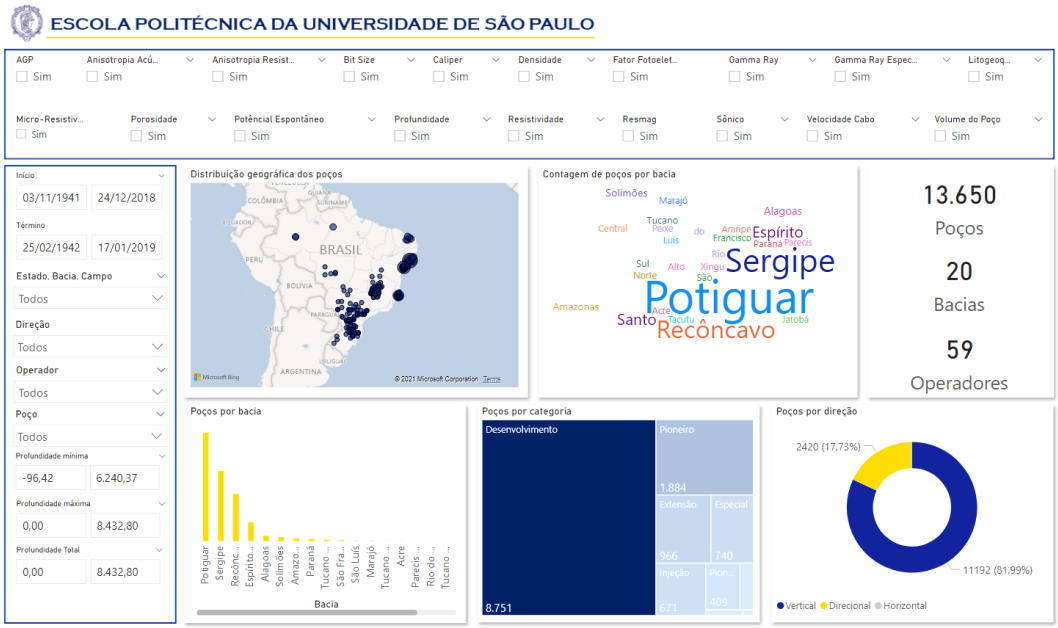


Figura 2 - Cenário de dashboard visando a busca por mnemônicos de interesse

Também é possível abordar um cenário (Figura 3) onde o usuário tenha a opção de pesquisar por um nome de poço específico, categoria de poço ou até mesmo bacia ou operador. Neste caso, é possível restringir inúmeras variáveis e verificar diversas condições específicas de análises de perfilagem de poços das bacias continentais brasileiras ou até mesmo de outras bases de dados originais e inéditas.

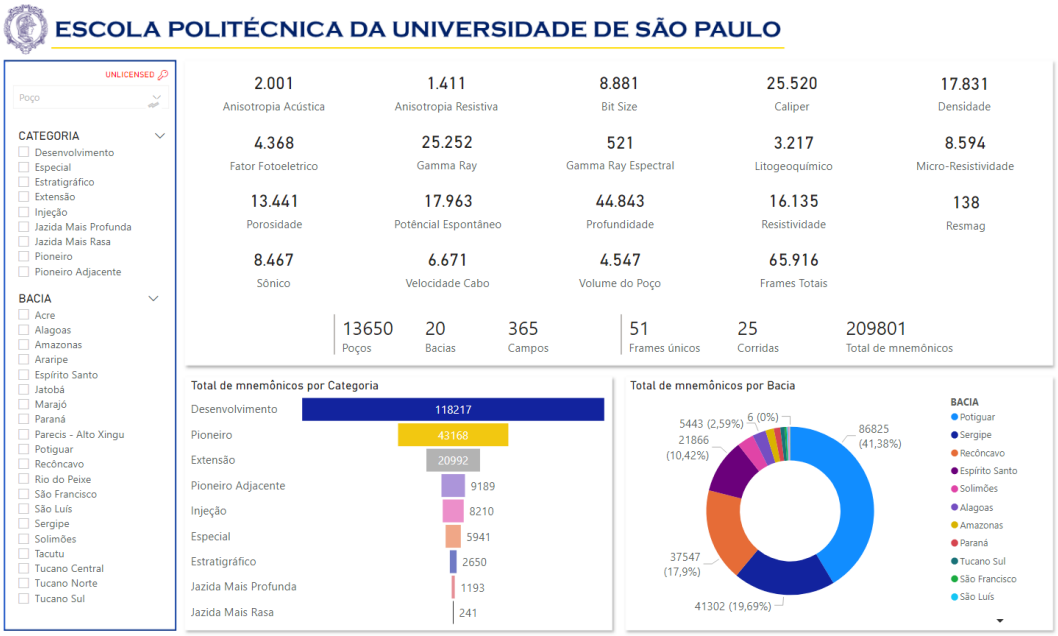


Figura 3 - Cenário de dashboard visando a busca por especificidade e gama de mnemônicos

4. Conclusão

O projeto foi composto por três pilares básicos, a saber:

i. Código indexador do acervo de dados para transformá-los em um banco não-relacional, no formato .JSON estará disponível no acervo da disciplina TCC 2 da Escola Politécnica.

ii. Análise, categorização e agrupamento de mnemônicos, que visou garantir uma factibilidade ao trabalho fazendo todo o processo de triagem dos mnemônicos de interesse nesse primeiro momento de pesquisa;

iii. Ferramenta de visualização em dashboards e dicionários, facilitando a futura interpretação dos dados e metadados antes compilados em arquivos dos mais diversos formatos em pastas e subpastas.

A reunião destes três pilares possibilitou a otimização do processo de busca por poços e suas respectivas assembleias de perfis. O trabalho desenvolvido favorece buscas simples e eficientes, o que propicia a redução de hora humana trabalhada em processos simples de localização de perfis associados aos poços, sejam estes exploratórios ou produtores.

4.1. Contribuições e trabalhos futuros

Este trabalho dará suporte a futuras pesquisas e consultas à base de dados de poços onshore brasileiros, bem como introduz um conceito simples que pode ser expandido também aos poços das bacias offshore. A rápida e precisa triagem e coleta de metadados dentro de inúmeras bases de dados de arquivos de poços nos mais diversos formatos permite transportar essas informações para um ambiente de busca e visualização gráfica, como por exemplo através da geração de novos dashboards e apresentações.

No perímetro de visualização de dados é simples e factível agregar conteúdos de plotagem interativa, clusterização e ambientes dashboard das mais diversas formas para auxiliar na interpretação das informações contidas em uma base original já processada, criando um panorama geral dos dados catalogados através dessas ferramentas em eventuais consultas de terceiros.

5. Referências

AGÊNCIA NACIONAL DO PETRÓLEO, GÁS NATURAL E BIOCOMBUSTÍVEIS - SUPERINTENDÊNCIA DE DADOS TÉCNICOS – SDT. **Nota Técnica nº 073/2016/SDT.** 2016.

ANGWIN, M.; NELSM, J. L.; SYRETT, B. C. 1996. **Technical Database Design.** The NACE International Annual Conference and Exposition. 1996.

ANP. **Acervo de Dados.** Disponível em: <http://www.anp.gov.br/exploracao-e-producao-de-oleo-e-gas/dados-tecnicos/acervo-de-dados>. Acesso em 22 de jul. de 2021.

CODD, E. F. A **Relational Model of Data for Large Shared Data Banks.** Communications of the ACM, v. 13, n. 6, p. 377–387, 1 jun. 1970.

COX, N. **Exploratory Data Mining and Data Cleaning.** Journal of Statistical Software, v. 11, 27 out. 2004.

Data B Inteligência. **O que é Power BI e por que usar em seu negócio?** Disponível em: <https://databinteligencia.com.br/o-que-e-power-bi-e-por-que-usar-em-seu-negocio/>. Acesso em 21 de nov. de 2021

DE ANDRADE, P. A. **Assinatura de Variáveis em Perfis de Poços na determinação de zonas de interesse para óleo e gás com base em Mapas Auto-Oganizáveis (SOM)**. 2019.

ENDEL, F.; PIRINGER, H. **Data Wrangling: Making data useful again.** IFAC-PapersOnLine. Elsevier, 1 fev. 2015

FRACZEK, K.; PLECHAWSKA-WOJCIK, M. Comparative Analysis of Relational and Non-relational Databases in the Context of Performance in Web Applications. 153-164. 10.1007/978-3-319-58274-0_13. 2017.

Informática UFSC. **Conceito Básicos da Teoria de Grafos.** Disponível em:< https://www.inf.ufsc.br/grafos/definicoes/definicao.html>. Acesso em 21 de nov. de 2021.

JATIN; BATRA, S. **MONGODB Versus SQL: A Case Study on Electricity Data**. p. 297–308. 2016

KANDEL, S. et al. Research directions in data wrangling: Visualizations and transformations for usable and credible data. Information Visualization, v. 10, n. 4, p. 271–288, 2 out. 2011.

Microsoft Docs. **Noções básicas de dashboard.** Disponível em: <https://docs.microsoft.com/pt-br/power-bi/create-reports/service-dashboards>. Acesso em 19 de nov. de 2021.

PHIRI, H.; KUNDA, D. **A Comparative Study of NoSQL and Relational Database.** Zambia ICT Journal, v. 1, p. 1–4, 27 dez. 2017.

Schlumberger. **Curve Mnemonic Dictionary.** Disponível em: <https://www.apps.slb.com/cmd/index.aspx> Acesso em 19 de nov. de 2021.

WICKHAM, H. **Tidy data.** Journal of Statistical Software, v. 59, n. 10, p. 1–23, 1 set. 2014.