

# 9º CONGRESSO BRASILEIRO DE PESQUISA E DESENVOLVIMENTO EM PETRÓLEO E GÁS



**9º CONGRESSO**  
Brasileiro de P&D em  
**PETRÓLEO E GÁS**

**Maceió, AL**  
de 09 a 11 de novembro  
**2017**

Realização:



**ABPG**  
ASSOCIAÇÃO BRASILEIRA DE P&D EM  
PETRÓLEO E GÁS



## **TÍTULO DO TRABALHO:**

DATA SCIENCE FOR OIL: ESTADO DA ARTE E PROPOSTA DE FERRAMENTA COMPUTACIONAL BASEADA EM CIÊNCIA DE DADOS PARA INDÚSTRIA DE PETRÓLEO E GÁS

## **AUTORES:**

Gustavo Coelho Lopes, Eduardo Setton Sampaio da Silveira, Josué Domingos da Silva Neto, Heitor Soares Ramos Filho

## **INSTITUIÇÃO:**

Engenharia de Petróleo, Centro de Tecnologia, Universidade Federal de Alagoas, josuedsneto@gmail.com, 2Engenharia Civil, Centro de Tecnologia, Universidade Federal de Alagoas

*Este Trabalho foi preparado para apresentação no 8º Congresso Brasileiro de Pesquisa e Desenvolvimento em Petróleo e Gás - 9º PDPETRO, realizado pela Associação Brasileira de P&D em Petróleo e Gás - ABPG, no período de 09 a 11 de novembro de 2017, em Maceió/AL. Esse Trabalho foi selecionado pelo Comitê Científico do evento para apresentação, seguindo as informações contidas no documento submetido pelo(s) autor(es). O conteúdo do Trabalho, como apresentado, não foi revisado pela ABPG. Os organizadores não irão traduzir ou corrigir os textos recebidos. O material conforme, apresentado, não necessariamente reflete as opiniões da Associação Brasileira de P&D em Petróleo e Gás. O(s) autor(es) tem conhecimento e aprovação de que este Trabalho seja publicado nos Anais do 9º PDPETRO.*

## **DATA SCIENCE FOR OIL: ESTADO DA ARTE E PROPOSTA DE FERRAMENTA COMPUTACIONAL BASEADA EM CIÊNCIA DE DADOS PARA INDÚSTRIA DE PETRÓLEO E GÁS**

### **Abstract**

The stages of exploitation, development and production of oil and gas produce a great volume of data (Spath, 2014). Data science is the interdisciplinary field responsible for extracting knowledge from data using principles and techniques of computing, engineering, statistics, economics, among other disciplines, to create innovative solutions through automatic data analysis. This data, that can be structured, non-structured or even semi-structured, is not yet broadly used, specifically because of the challenge that is deciding which data is relevant.

Big Data applications are in its early developments in the oil and gas industry, however, according to Spath (2014), with the increase of field and well monitoring, specially due to the pressure being made by regulatory agencies over the world, data analysis is the next great technology innovation in the oil industry. Mehta (2016) cites that data analysis is one of the great priorities for oil and gas companies in 2018.

In Brazil, data regarding the oil industry is under the responsibility of the Nacional Oil Agency (ANP) and can be found in reports, tables and bulletins, that way it lacks a centralized database with intuitive access. Beyond ANP, there are reports that can be found in initiatives brought by organizations such as OPEC, BP, EIA, the initiatives are JODI, Knoema, Dataviva or even in commercial softwares as the RystadEnergy. One of the problems found in the industry and research centers is related to the time-consuming task of finding, obtaining and aggregating the data for further analysis. In an environment where decision making has to be near instantaneous, time is a crucial matter.

This work presents the analysis of the development of a computational database for oil and gas – Data Science for Oil, which aims at obtaining, treating and analyzing the data that is broadly available in the internet and the databases of regulatory agencies and oil and gas companies to provide a complete and integrated database for oil and natural gas in Brazil.

### **Introdução**

Estamos na quarta revolução industrial, Wang et al. (2016), Bloem et al. (2014), e os dados estão aqui para ficar. Só resta às indústrias tradicionais aceitarem essa nova revolução ou arriscarem serem consideradas atrasadas.

Na era da informação, o modo mais confiável de se tomar decisões se tornou o Big Data (Lohr, 2012), que se trata do tratamento de dados que possuem certas características em comum, sendo as principais, o Volume, a Variedade e a Velocidade na qual essas informações se tornam disponíveis. Nessa nova forma de tratar os dados faz-se uma análise histórica e em tempo real para auxílio da tomada de decisão, o período de tomar decisões baseadas em instinto está com seus dias contados.

O volume de dados tratados na indústria do óleo e gás sempre foi historicamente grande, dados contidos em data centers de empresas modernas de óleo e gás podem chegar à grandeza de 20 petabytes (Beckwith, 2011), principalmente agora com o advento da Internet das Coisas e sensores com informação em tempo real. Esses dados sempre foram usados para auxiliar a tomada de decisão, porém em uma menor escala. Não há o uso sistemático e contínuo dos dados para tomar decisões mais fundadas (Perrons, 2015).

As agências reguladoras e alguns grandes participantes da indústria ao redor do mundo consideram informação de grande importância a presença desses dados, tais como a Organização dos Países Produtores de Petróleo (OPEP), a Internacional Energy Agency (IEA), a Agência Nacional de Petróleo (ANP), a BP, a Shell, entre outros. Essas empresas e instituições geralmente disponibilizam os dados em seus sites, que podem ser estruturados, não estruturados ou semi-estruturados (Hamzeh, 2016).

A indústria do petróleo ainda não está preparada para essa mudança no paradigma, pois a sua estrutura ainda não é compatível com a ideal para a adoção desse tipo de tecnologia. A mudança na indústria deve ocorrer gradualmente e com um objetivo claro e alcançável (Lavallo, 2011). No Brasil, não há uma plataforma que integre esses dados e alavanque a indústria nessa direção. Por isso, esse trabalho se propõe a estudar sobre o assunto e mostrar se esse tipo de ferramenta é ou não necessária e, caso seja, se exige esforços muitos grandes.

### **Metodologia**

O trabalho realizado foi feito no formato de revisão bibliográfica e de prospecção tecnológica. Usados como fontes de dados científicos o Web of Science, o Scielo, Google Acadêmico. Como fonte para prospecção tecnológica foram utilizadas bases de dados de patentes e bases de dados de software, como o GitHub. Pelo escopo do trabalho, também foram feitas buscas de empresas que realizam esse tipo de software de visualização e ferramentas de Big Data.

### **Resultados e Discussão**

Na prospecção tecnológica e de mercado foram encontrados diversos atores chave para a realização da ferramenta proposta, tais como:

- **Dataviva:** Ferramenta de Big Data realizada pela empresa Datawheel em parceria com a FAPEMIG e o Governo de Minas Gerais. O Dataviva integra bases de dados de diversos ministérios brasileiros e os apresenta de forma interativa e de fácil compreensão.
- **Datawheel:** Empresa que produz softwares de Big Data, seus principais clientes são entidades de governo e seus principais produtos são o Dataviva, o Datausa e o Observatory for Economic Complexity. A empresa também possui um portal que fornece software livre de visualização de dados, o D3Plus.
- **JODI:** JODI é uma iniciativa de um conjunto de organizações não governamentais, tais como a Organização dos Países Exportadores de Petróleo (OPEP), International Energy Agency (IEA), a divisão estatística das nações unidas, entre outros. O JODI Oil oferece dados de petróleo, como produção, consumo e exportação de vários países do mundo. E o JODI Gas oferece o mesmo tipo de dado para o setor de gás.
- **BP:** A BP fornece anuários estatísticos anuais de produção de energia no mundo. As bases de dados abrangem tipos de energia além do petróleo, como hidroelétrica, solar, carvão e gás.

Esses atores fornecem uma boa fundação para a realização da ferramenta proposta, de modo que resta apenas a fundamentação científica para a construção de forma a satisfazer as necessidades dos usuários. Desse modo, foi feita uma revisão bibliográfica de como a ciência de dados e o Big Data geram valor para usuários e auxiliam o processo de tomada de decisão, bem como o tipo de entidade que usa esse tipo de ferramenta e como estimular esse uso.

De acordo com Miller e Mork (2013), existem diversas etapas na cadeia de criação de valor do Big Data, basicamente divididos entre Descoberta, Integração e Exploração, de acordo com a figura 1.

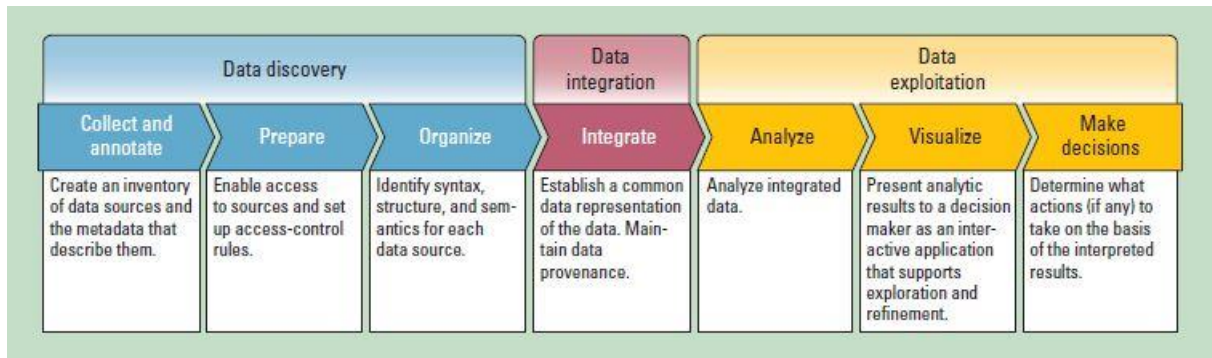


Figura 1: Cadeia de criação de valor do Big Data

Fonte: Miller e Monk, 2013 p. 58

Uma boa ferramenta de visualização de dados deve ser simples, de fácil interpretação e acesso, dois exemplos são o JODI Oil (Figura 2) e o Dataviva (Figura 3). O JODI Oil apresenta os dados na fase de Organização dos dados, já o Dataviva apresenta na fase de Visualização. Pode-se claramente diferenciar o valor dos dados apresentados pelas plataformas, o dado do Dataviva possui um valor superior ao JODI Oil.

Reports

Table

Chart

</

Figura 2: Apresentação de dados JODI Oil

Fonte: JODI Oil, acessado em Ago, 2017



### Perfil Geral

O curso/campo de Direito é o 1º em número de matrículas no Brasil. A universidade que possui mais alunos é Universidade Estácio de Sá, com 25,6 mil estudantes. No total, o curso possui 813 mil alunos matriculados em todo o país.

#### MATRÍCULAS (2015)

Principal Universidade por Número de Alunos Matriculados	UNIVERSIDADE ESTÁCIO DE SÁ 29,5 Mil
Principal Município por Número de Alunos Matriculados	SÃO PAULO 63,9 Mil
Principal Universidade por Número de Alunos Ingressantes	UNIVERSIDADE ESTÁCIO DE SÁ 10,2 Mil

Figura 3 Apresentação de dados do Dataviva

Fonte: Dataviva, acessado em Ago, 2017

Pode-se observar que ambas as plataformas possuem dados estruturados, em formato de fácil manipulação e ambas disponibilizam seus dados para serem importados nos formatos .xls e .csv, formatos ideais para o uso de ferramentas de visualização. Porém, o Dataviva é muito mais acessível e apresenta as informações já analisadas e apresenta conclusões importantes logo ao acessar o dado. O JODI Oil apresenta o dado em formato bruto, tornando o processo de visualização e análise não tão elementar.

A ferramenta proposta deverá obter dados de diversas fontes, tais como, bases de dados do governo, que englobam os dados da ANP e os Balanços energéticos nacionais, dados de empresas privadas, como a BP, dados de organizações internacionais como EIA e JODI Oil.

Com esses dados deverão ser feitas referências cruzadas para analisar a consistência dos dados de acordo com as diversas fontes, então os dados serão preparados, organizados e integrados, utilizando de ferramentas como Hadoop Infrastructure, Microsoft MURA e IBM PureData. E por fim esses dados serão analisados e dispostos à visualização utilizando-se do Python e bibliotecas com o D3 Plus.

As últimas etapas deverão usufruir dos princípios do storytelling, que pertinem que os dados contem sua própria história, pois para cada dado existe um tipo de visualização ideal (Krzywinski, 2013)

Ferramentas de Big Data podem ser de diversas magnitudes e em diversos estágios de desenvolvimento, no caso estudado existem vários dados amplamente disponíveis em diversos tipos de maturação, tais quais o JODI Oil, que possui dados desde 2002 dos países da OPEP, ou a BP, que possui dados diversos de energia no mundo desde 1951.

Porém, Big Data não acaba na coleta de dados, os dados devem ser preparados e armazenados, preferencialmente em nuvem, por serviços como Amazon Web Service, então devem ser organizados em infraestrutura de tecnologia da informação que seja robusta como a base de dados.



Por fim os dados devem ser analisados de forma a serem inteligíveis e de fácil acesso para os tomadores de decisão.

## Conclusões

De acordo com o estudado nesse trabalho, conclui-se que não há ferramenta nacional para visualização de dados referentes à indústria de óleo e gás, apesar da ampla disponibilidade de dados que tornam possível a criação de tal.

Ademais, vale enfatizar a importância desse tipo de ferramenta como um avanço no processo de tomada de decisão para gestores e membros da indústria. Portanto a criação dessa ferramenta se torna imprescindível e necessária para o ingresso da indústria nacional na quarta revolução industrial.

## Agradecimentos (Tamanho: 11 – Fonte: verdana)

Texto-Tamanho: 11 – Fonte: Times New Roman

## Referências Bibliográficas

1. WANG, Shiyong et al. Towards smart factory for industry 4.0: a self-organized multi-agent system with big data based feedback and coordination. **Computer Networks**, [s.l.], v. 101, p.158-168, jun. 2016. Elsevier BV. <http://dx.doi.org/10.1016/j.comnet.2015.12.017>.
2. BLOEM, Jaap, et al. **The Fourth Industrial Revolution**. *Things Tighten*, 2014.
3. LOHR, Steve. **The age of big data**. *New York Times*, 2012, 11.2012.
4. BECKWITH, Robin. Managing Big Data: Cloud Computing and Co-Location Centers. **Journal Of Petroleum Technology**, [s.l.], v. 63, n. 10, p.42-45, 1 out. 2011. Society of Petroleum Engineers (SPE). <http://dx.doi.org/10.2118/1011-0042-jpt>.
5. PERRONS, Robert K.; JENSEN, Jesse W.. Data as an asset: What the oil and gas sector can learn from other industries about “Big Data”. **Energy Policy**, [s.l.], v. 81, p.117-121, jun. 2015. Elsevier BV. <http://dx.doi.org/10.1016/j.enpol.2015.02.020>.
6. Hamzeh, Hamed. **Application of Big Data in Petroleum Industry**, fev. 2016.
7. LAVALLE, Steve, et al. Big data, analytics and the path from insights to value. *MIT sloan management review*, 2011, 52.2: 21.
8. MILLER, H. Gilbert; MORK, Peter. From Data to Decisions: A Value Chain for Big Data. **It Professional**, [s.l.], v. 15, n. 1, p.57-59, jan. 2013. Institute of Electrical and Electronics Engineers (IEEE). <http://dx.doi.org/10.1109/mitp.2013.11>.
9. KRZYWINSKI, Martin; CAIRO, Alberto. Points of view: Storytelling. **Nature Methods**, [s.l.], v. 10, n. 8, p.687-687, 30 jul. 2013. Springer Nature. <http://dx.doi.org/10.1038/nmeth.2571>.
10. DATAWHEEL (Estados Unidos). **Datawheel**. Disponível em: <<http://www.datawheel.us/>>. Acesso em: 15 ago. 2017.
11. DATAVIVA (Minas Gerais). **Dataviva**. Disponível em: <<http://www.dataviva.info>>. Acesso em: 15 ago. 2017.
12. JODI. **JODI**. Disponível em: <<https://www.jodidata.org/>>. Acesso em: 15 ago. 2017.
13. DATA USA. **DATA USA**. Disponível em: <<https://datausa.io/>>. Acesso em: 15 ago. 2017.