VDS – Novo formato para dados sísmicos da Petrobras

Francisco Aquino, Marcelo Landini, Pascoal Melo, Victor Portavales, Petrobras – Exploração

Copyright 2024, SBGf - Sociedade Brasileira de Geofísica

Este texto foi preparado para a apresentação no IX Simpósio Brasileiro de Geofísica, Salvador, 08 a 10 de outubro de 2024. Seu conteúdo foi revisado pelo Comitê Técnico do X SimBGf, mas não necessariamente representa a opinião da SBGf ou de seus associados. É proibida a reprodução total ou parcial deste material para propósitos comerciais sem prévia autorização da SBGf.

Resumo

O Formato SEG-Y foi criado em 1975 e desde então vem sendo utilizado pela indústria de petróleo como formado de representação de dados sísmicos, passando por poucas modificações e atualizações até os dias de hoje. A Petrobras adota o formato SEG-Y (rev.0) como o formato padrão de recebimento e armazenamento de dados sísmicos desde 2002.

Por se tratar de um formato de dados de acesso sequencial e com organização interna pré-definida (ex. organização por sequência de tiros), apresenta limitações e restrições ao uso direto em fluxos de trabalho computacionalmente intensivos e ou de leitura e escrita intensivos.

Recentemente, com o início da popularização dos ambientes de computação em nuvem com uso de tecnologia de armazenamento do tipo "Object Storage" e do crescente uso de técnicas de aprendizado de máquina (ex. Machine Learning e Deep Learning), tornou-se necessária a utilização de um novo formato de representação de dados sísmicos.

Em 2018 o Open Group[1] criou o consórcio OSDU (Open Subsurface Data Universe), formado pelos principais operadores de Petróleo, empresas de serviços de E&P e empresas de tecnologia de nuvem, para a definição e implantação, de uma plataforma de dados aberta, padronizada e nativa para ambientes de computação em nuvem, para uso pelo segmento de energia, visando a eliminação dos silos de dados e permitindo a ampla interoperabilidade de dados entre diferentes aplicações.

O OSDU adotou como um dos seus padrões de representação de dados sísmicos o formato VDS (Volume Data Store), denominado internamente pelo consórcio OSDU como OpenVDS, formato este totalmente adequado para as necessidades de fluxos de trabalhos computacionalmente intensivos no ambiente de computação em nuvem.

Este trabalho apresenta o formato VDS, os benefícios de seu uso em substituição ao formato SEG-Y e o modelo de adoção deste formato pela Petrobras, como formato padrão de dados sísmicos em ambiente de computação em nuvem em conjunto com a adoção da Plataforma de dados OSDU.

Introdução

O Formato SEG-Y foi criado em 1975 pela SEG (Society of Exploration Geophysicists) como padrão de representação de dados sísmicos para armazenamento em disco, como uma evolução as famílias de formatos de representação de dados sísmicos para fitas magnéticas (SEG-A, SEG-B, SEG-C e SEG-D).

O formato SEG-Y atualmente é amplamente utilizado pelo mercado de Petróleo e de Geofísica como o padrão de representação e intercambio de dados sísmicos brutos e processados, podendo ser utilizado tanto em fitas magnéticas como em discos com tecnologia do tipo "File Store". A Petrobras adota o formato SEG-Y (Rev. 0) como formato padrão único para recebimento e envio de dados sísmicos brutos e processados desde 2002, sendo que desde 2015 só recebe os arquivos SEG-Y através do envio direto destes arquivos através de uma rede dados com canais de fibra ótica e internet (RDG – Rede de dados Geofísicos).

Devido a sua estrutura interna (Figura 1), o padrão SEG-Y apresenta algumas características que limitam sua performance para uso em fluxos técnicos computacionalmente intensivos, tais como: (i) O acesso sequencial aos dados, que impõe a necessidade de leitura de todo os dados; (ii) a necessidade de uma organização prévia dos dados de acordo como uma sequencia no momento de sua geração, que impõe a necessidade de leitura de todo dado, seguida de uma operação reordenação (Sort) para utilização do dado em um uma nova sequência e (iii) a limitação de estrutura para a inclusão de metadados que descrevam corretamente todos os processos técnicos que o dado geofísico foi submetido e limitam o seu descobrimento posterior.

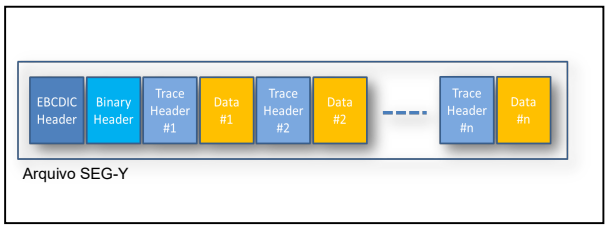

Figura 1 – Estrutura interna de um arquivo SEG-Y (Rev. 0).

Ao longo dos anos a indústria de Petróleo vem contornando estas limitações através da adoção do formato SEG-Y apenas como padrão de armazenamento e troca de dados e usando a conversão do formato SEG-Y para diversos e diferentes formatos proprietários de representação de dados sísmicos, mais adequados para

[1] Open Group – Consórcio Global de empresas para definição de padrões abertos de tecnologia e serviços

uso nas aplicações de processamento e análise (interpretação), contribuindo para a geração de um crescente número silos de dados proprietários e também para com o consequente aumento das dificuldades de descobrimento dos dados existentes e da interoperabilidade de dados entre diferentes aplicações.

Com o início do uso da computação em nuvem pelas empresas de Petróleo, com o uso das tecnologias de armazenamento do tipo “Object Store” (armazenamento padrão do ambiente de nuvem), e de técnicas de aprendizado de máquina (Machine Learning) e em especial as técnicas de aprendizado profundo (Deep Learning), tornou necessário a adoção de um novo formato para representação dos dados sísmicos, que potencializa-se tanto o uso das tecnologias disponíveis no ambiente de nuvem como a padronização e a interoperabilidade de dados.

Em 2018 a Open Group, consórcio global formado por empresas de tecnologia que promove a criação e adoção de padrões abertos para produtos e serviços de tecnologia da informação, criou um consórcio específico para a área de energia, denominado OSDU (Open Subsurface Data Universe), com o objetivo de criar uma plataforma de dados em ambiente de nuvem, aberta e padronizada, de forma a otimizar e potencializar o descobrimento, acesso e o uso dos dados no ambiente de computação em nuvem. O consórcio OSDU é formado pelas principais empresas operadoras de Petróleo (Shell, BP, Chevron, Total, Equinor, ExxoMobil, Petrobras, ...), por empresas de serviços de G&G (CGG, PGS, TGS, Halliburton, Shclumberger, Aspen Tech, Bluware, ...) e de tecnologia de Nuvem (AWS, Microsoft e Google).

O Consórcio OSDU definiu que suportaria tanto o formato SEG-Y quanto o formato VDS (Volume Data Store), denominado internamente como OpenVDS, como formatos de representação para dados sísmicos, abrindo assim a oportunidade de utilização VDS com formato padrão de representação de dados sísmicos com características mais adequadas as necessidades de performance e armazenamento, exigidas pelos processos de trabalho computacionalmente intensivos e pelos processos de aprendizado de máquina (machine Learning) e de aprendizado profundo (Deep Learning).

Em 2022 a Petrobras conduziu um estudo para avaliar os benefícios, vantagens e o modelo de adoção do formato VDS, como o formato padrão de representação de dados sísmicos em ambiente de nuvem incorporado a sua implantação do OSDU, bem como as alterações e modificações necessárias no atual processo de Gestão de dados sísmicos adotado na Petrobras.

O formato VDS

O formato VDS foi criado em 2000 pela empresa Bluware. Em 2020 a empresa dou o código aberto do formato VDS para o consórcio OSDU, sendo denominado internamente no OSDU como OpenDVS.

O VDS é um formato para armazenamento de dados com acesso aleatório, rápido para dados volumétricos multidimensionais, suportando até 6 dimensões distintas num mesmo arquivo, sendo capaz de prover taxas de leitura e escrita eficientes para dados da ordem de PetaBytes.

Os dados sísmicos são representados no VDS na forma de cubos 3D (Bricks) de 128 x 128 x 128 amostras ou 64 x 64 x 64 amostras, nos formatos 64 ou 32 Bit Floating Point ou 8,16,32 e 64 Bit Integer ou 2 ou 4 components Vectors, com suporte ao uso ou não da compactação dos dados por diferentes algoritmos de compactação (figura 2).

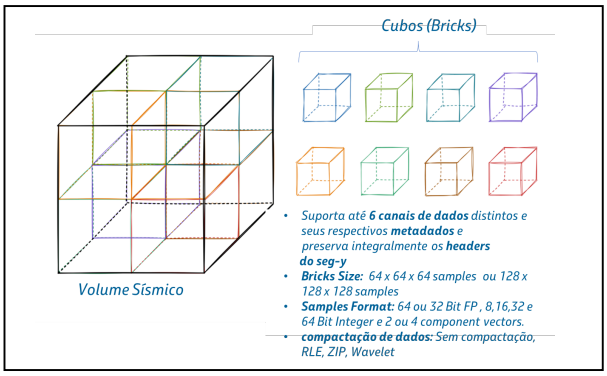

Figura 2 – Estrutura interna dos dados de um arquivo VDS.

Junto aos dados sísmicos também são armazenados os metadados que descrevem os dados, incluindo uma estrutura capaz de armazenar integralmente todos os headers do arquivo SEG-Y original do qual foi gerado o VDS, permitindo assim a recomposição total e idêntica do arquivo SEG-Y a partir do arquivo VDS gerado (Figura 3).

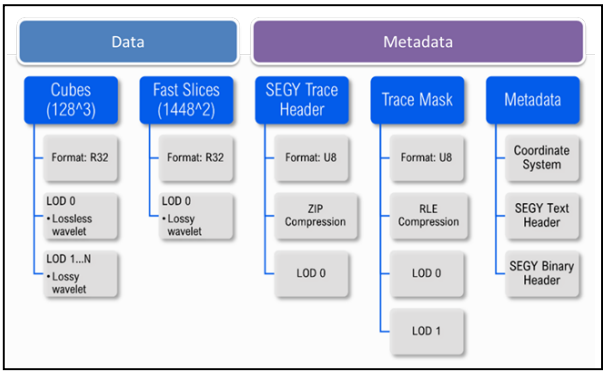

Figura 3 – Estrutura interna dos dados e metadados do formato VDS.

Adicionalmente aos cubos de dados, o VDS possui suporte a geração de dados sísmicos em seções 2D (Slices) em três dimensões e pode utilizar os recursos de multi-resolução (LOD – Level of Detail) com até 12 níveis de subamostragem do dado original, onde cada nível

possui 1/8 do número de amostras do nível anterior (figura 4).

Outra característica importante do VDS é o suporte ao uso de diferentes tecnologias de armazenamento, podendo um mesmo arquivo VDS ser utilizada na forma de um arquivo tradicional em dispositivos de armazenamento do tipo "File Storage", como em dispositivos de armazenamento do Tipo "Object Storage", tipo de armazenamento padrão nos ambientes de nuvem, mantendo integralmente as mesmas características técnicas em ambos os tipos de armazenamento. No caso do armazenamento tipo "Object Storage", cada cubo de dados do VDS é armazenado como um arquivo independente, podendo ser acesso diretamente pelas aplicações.

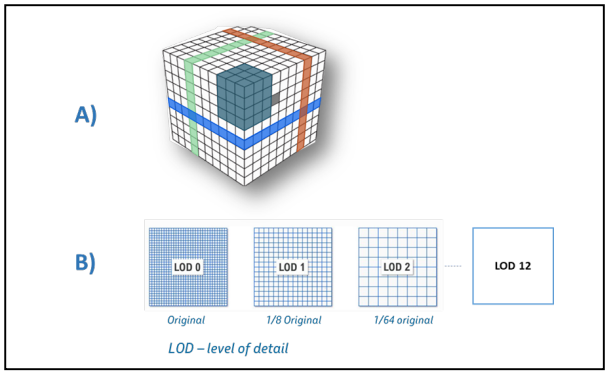

Figura 4 – A) Estrutura de armazenamento de Slices 2D, de até três eixos distintos (x,y e z) e B) Multi-Resolução do formato VDS (LOD).

Benefícios do uso do formato VDS

O estudo conduzido pela Petrobras demonstrou a existência de inúmeros benefícios para a adoção do formato VDS como formato de representação de dados sísmicos em ambiente computacional de nuvem, especialmente quando utilizado como parte integrante da plataforma OSDU, entre os quais podemos destacar:

1. Acesso aleatório aos dados sísmicos no formato VDS resulta em uma performance de acesso I/O e de processamento muito superior as obtidas como o formato SEG-Y;
2. O suporte do formato VDS ao uso de diferentes tecnologias de armazenamento (File Storage e Object Storage), de forma transparente a aplicação e sem perder suas características de acesso aleatório.
3. O suporte a funcionalidades para aceleração dos processos de visualização dos dados sísmicos, como a geração prévia de slides 2D em três diferentes direções (x – inline, y – crossline e z – tempo/profundidade) e a funcionalidade de multi-resolução, com a geração prévia de subamostragens do dado sísmico.
4. O suporte a diferentes tipos de compactação dos dados sísmicos que permitem a utilização de diferentes níveis de compactação de acordo com os requisitos de utilização do dado sísmico.
5. A potencialização da aplicação de técnicas de machine learning e Deep Learning, através da diminuição das etapas prévias de preparação e condicionamento do dado sísmicos, em virtude do acesso aleatório e da estrutura de Bricks do formato VDS, assim como pela diminuição do volume dos dados sísmicos a serem processados pelo uso da opção de compactação de dados (figura 5).
6. A adoção do formato VDS pelo consórcio OSDU, potencializando o uso do VDS pelo suporte ao acesso dos arquivos através do OSDU, que facilita o descobrimento, o acesso e a adoção de políticas de controle de acesso ao dado sísmico;

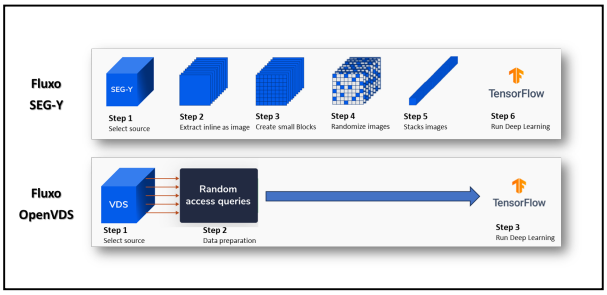

Figura 5 – Comparação entre as etapas de manipulação e condicionamento dos dados sísmicos para a aplicação de um processo de Deep Learning, para um dado sísmico no formato SEG-Y e VDS.

Resultados e Conclusões

Com base nos inúmeros benefícios do uso do formato VDS em substituição ao uso do formato SEG-Y, a Petrobras decidiu pela adoção do uso formato VDS, incorporado ao seu ambiente OSDU (OpenVDS) como o novo e único formato de representação de dados sísmicos empilhados em ambiente de nuvem.

Entretanto, visando manter a compatibilidade com os processos internos de gestão de dados, com as aplicações internas existentes e com as necessidades de troca de dados com empresas parceiras, de serviços e ANP, o formato SEG-Y será mantido como padrão de troca de dados com a Petrobras e como padrão de armazenamento de dados sísmicos em sua base de dados corporativa (BDIEP – Base de dados integrada de E&P).

Os arquivos sísmicos gerados e ou recebidos, após seu armazenamento na BDIEP, serão automaticamente convertidos para o formato VDS e carregados no OSDU em ambiente de nuvem, ficando disponíveis para acesso e consumo por todos os ambientes de análise, interpretação, tratamento e de ciência de dados disponíveis no ambiente de nuvem da Petrobras.

Visando potencializar o acesso e uso aos dados sísmicos no formato VDS definiu-se que para cada arquivos SEG-Y recebido serão gerados e carregados 2 arquivos VDS no ambiente OSDU: (1) um arquivo VDS sem qualquer tipo de compactação e com mutliresolução em todos os níveis (LOD 1 – 12), que deve ser utilizado em processos que exigem a melhor resolução numérica do dado (igual à do arquivo SEG-Y original) e (2) um arquivo VDS com uso de compactação da ordem de 5 vezes, com slices 2D em três direções e com mutli-resolução em todos os níveis (LOD 1 -12), que deve ser utilizado nos processos tradicionais de interpretação G&G e de fluxos de trabalho de aprendizado de máquina (Figura 6).

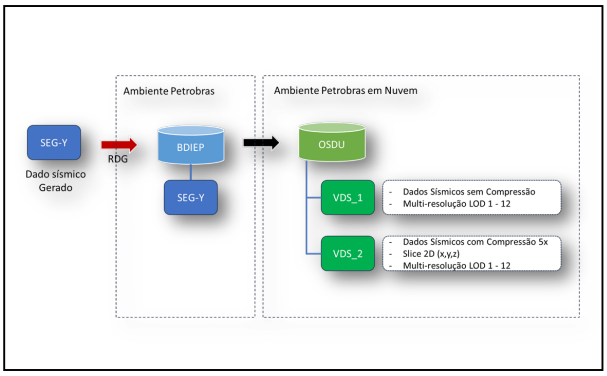

Figura 6 – Modelo de implantação do VDS na Petrobras, onde cada dado sísmico no formato SEG-Y recebido é armazenado na BDIEP e dois arquivos VDS são gerados e arquivos no OSDU.

Agradecimentos

Os autores deste trabalham agradecem a Petrobras pela autorização de sua publicação e em especial a todos os profissionais que trabalharam no estudo de avaliação do uso do formato VDS.

Referências

SEG-Y documentation, https://wiki.seg.org/wiki/SEG-Y

OSDU documentation, https://osdu.pages.opengroup.org/platform/domain-data-mgmt-services/seismic/open-vds/vds/specification/Introduction.html

Open Subsurface Data Universe Software / Data Definitions and Services / Data Definitions · GitLab (opengroup.org). (n.d.). Retrieved from https://community.opengroup.org/osdu/data/data-definitions

OpenVDS's documentation, opengroup.org: https://osdu.pages.opengroup.org/platform/domain-data-mgmt-services/seismic/open-vds/index.