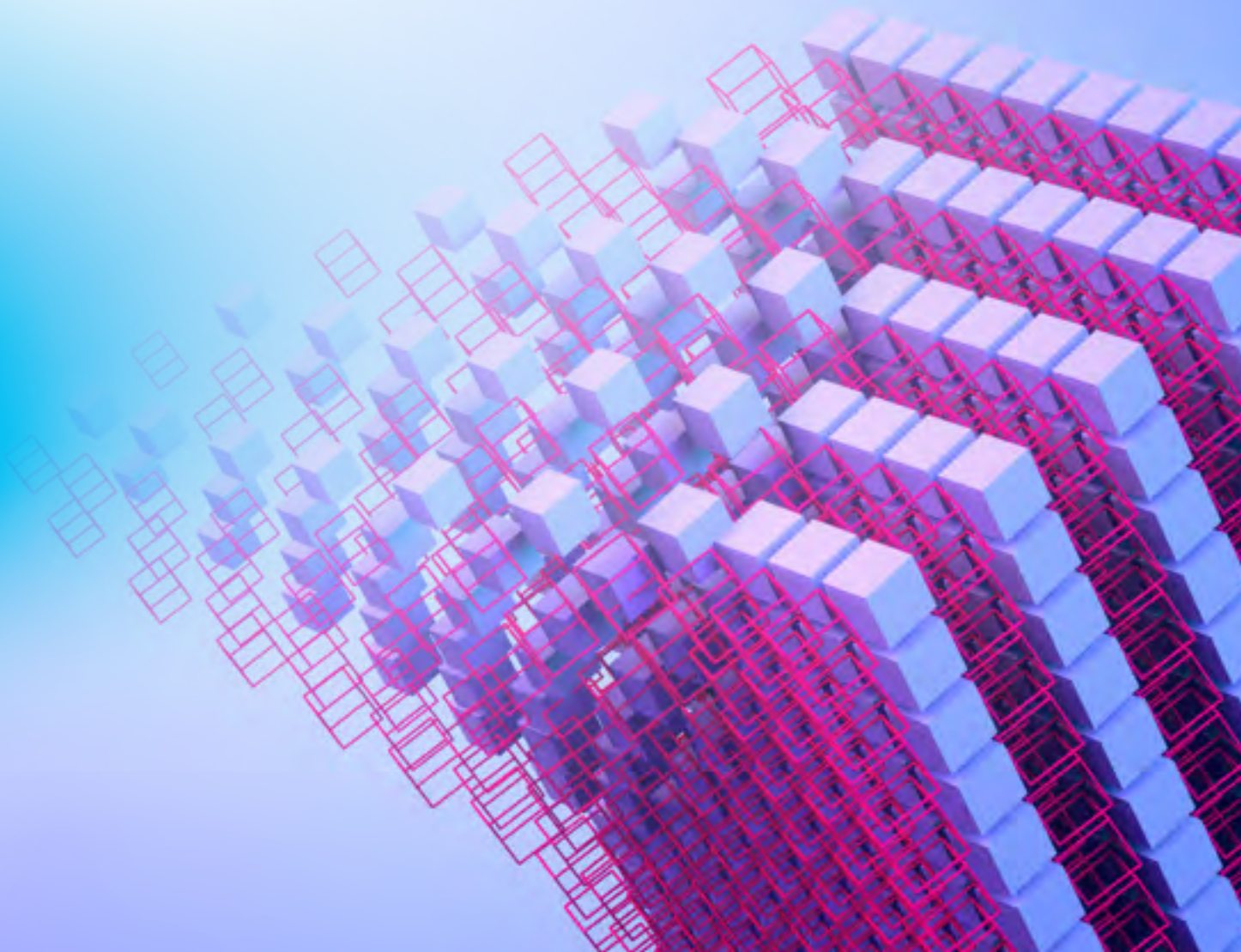
LTIMindtree

POINT OF VIEW

Migration of Seismic Data into Cloud hosted OSDU Data Platform

AUTHOR
Srinivasa Sudhakar Kadagala



Synopsis

The Oil and Gas (O&G) industry faces a massive challenge in handling vast volumes of subsurface data and adopting digital technologies for workflow enhancements. Leading Exploration and Production (E&P) companies and service providers have started adopting the Open Subsurface Data Universe (OSDU) data platform to store seismic data to address these challenges. The major challenge in deploying the OSDU data platform is migrating vast seismic data volumes to cloud-native data platforms and creating robust metadata information to properly index and identify the seismic volume.

Table of Contents

01 Introduction

In O&G companies, seismic data is one of the most valuable subsurface assets. Different companies have faced challenges in handling this vast data often stored in silos and unstructured formats. The unstructured data includes reports, documents, and spreadsheets in various languages and their relationship with the local file system is not in an electronically ready setup. Another big challenge is migrating huge volumes of seismic data into OSDU data platform simultaneously.

The sdutil is a command line utility tool developed by OSDU to upload and store seismic data in seismic store of OSDU data platform. The OSDU standardized APIs and Postman collections are used to convert the data format from SEGY to OpenVDS or OpenZGY in blob storage.

Many users face challenges in data transfer through sdutil process, which involves pre-requisite viz., virtual environment set up, configuration file (config.yaml) based on the different Cloud Service Providers (CSPs), authorization & environment variable setup along with the presence of the required files in the seismic-store-sdutil-master file.

Additionally, challenges exist in data format conversion, data quality assurance, data security when dealing with large volumes of seismic data. To overcome these challenges, many tools & techniques are used viz., data compression, encryption algorithms and optimizing the various processes involved for better efficiency and security of the data.

02 Adoption of the latest technologies

The latest developments in technology in recent times, like the evolution of cloud computing, big data, Artificial Intelligence (AI), and Machine Learning (ML), led many E&P and service companies to opt for the digital transformation of subsurface data, especially the upstream sector as per the customer's need. In the 1940s, scientists started exploring artificial brains, which has led to recent developments in AI, ML, big data, and cloud computing in terms of security, data governance, quick access, etc.

This made O&G operators, oilfield service companies, cloud service providers, and application developers collaborate with each other to form a centralized hub as a single reference data platform. They started working together to form a cloud-based platform to migrate data from various sources instead of storing data in siloes. Initially, companies like Shell and Schlumberger developed their cloud-based Subsurface Data Hubs, namely Shell Data Universe (SDU) and DelFi, respectively. Later, this led to the outcome of the Open Subsurface Data Universe (OSDU) Forum by The Open Group.

03 Standardized storage format on the cloud

During the adoption of digital transformation, all O&G companies had a big challenge of migrating and storing vast seismic data of different formats into a cloud-hosted OSDU data platform. Open Volume Data Storage (OpenVDS) and OpenZGY were introduced to the OSDU Data Platform to overcome this difficulty.

A OpenVDS on OSDU Data Platform

To avoid integration issues like having different storage formats on a cloud-based platform, a standardized data format, OpenVDS was adopted on the OSDU platform. In the absence of OpenVDS applications on the cloud, it is very difficult to store a non-standardized data format.

The Bluware Corporation developed OpenVDS, and it was contributed to the OSDU Data Platform for storing any kind of data in a single industry-standardized data format. OpenVDS is an open-source API consisting of a set of tools to read and write data in VDS format. It has powerful capabilities for accessing the data quickly, securely and visualize the seismic volumetric data interactively. Storing data in VDS is cost-effective and brings the customer business values. Open VDS is commonly known for seismic data storage, and the capabilities are classified as reading, storing, and writing in the OSDU Data Platform.

B OpenZGY on OSDU Data Platform

OpenZGY is another format available in the OSDU Data Platform as an open-source library used for storing seismic data. Generally, the SEGY format is converted to ZGY format, which is used in Schlumberger's proprietary Petrel software for seismic interpretation workflows. OpenZGY library also enables the creation of compressed OpenZGY files resulting in the reduction of storage space in OSDU, and can easily be accessible in other applications using OpenZGY library.

C Seismic data in blob storage on the cloud

Initially, many companies faced issues in ingesting bulk amounts of seismic data into OSDU and visualizing the seismic volumes on workstations. After Bluware Corporation made VDS open source and made it available on the OSDU Data Platform for all, it became easy for all O&G companies to migrate seismic data and store it in VDS format.

The benefits of having VDS on the OSDU Data Platform enable quick access and better streaming of seismic volumes. In other words, the VDS format stores signal data in bricks and slices as sample points, enabling the end user to access the data quickly. When visualizing seismic volumetric data on various E&P applications, the data imported is in decimated format with a low resolution, occupying less space than its input data. The bricks and slices of VDS are identical to bricks and slices stored in cloud-based object storage, making it easy to move the data from each other smoothly without any issues. Object storage, like Blob storage in the cloud, such as Platform as a Service (PaaS), makes it easier to store massive amounts of unstructured data (human or machine-generated data) due to its features like flexibility and elasticity in nature.

This led many major and minor O&G companies to move towards digital transformation and to leverage OpenVDS by storing bulk volumes of data; particularly the seismic data in petabytes and their metadata in VDS format in an object storage cloud service such as Microsoft Azure Blob storage, Google Cloud Storage, or Amazon Simple Storage Service (S3). Storing in OpenVDS format on a cloud will enable an end user to instantly access several types of data, from 2-dimensional to 6-dimensional seismic data (Figure 1).

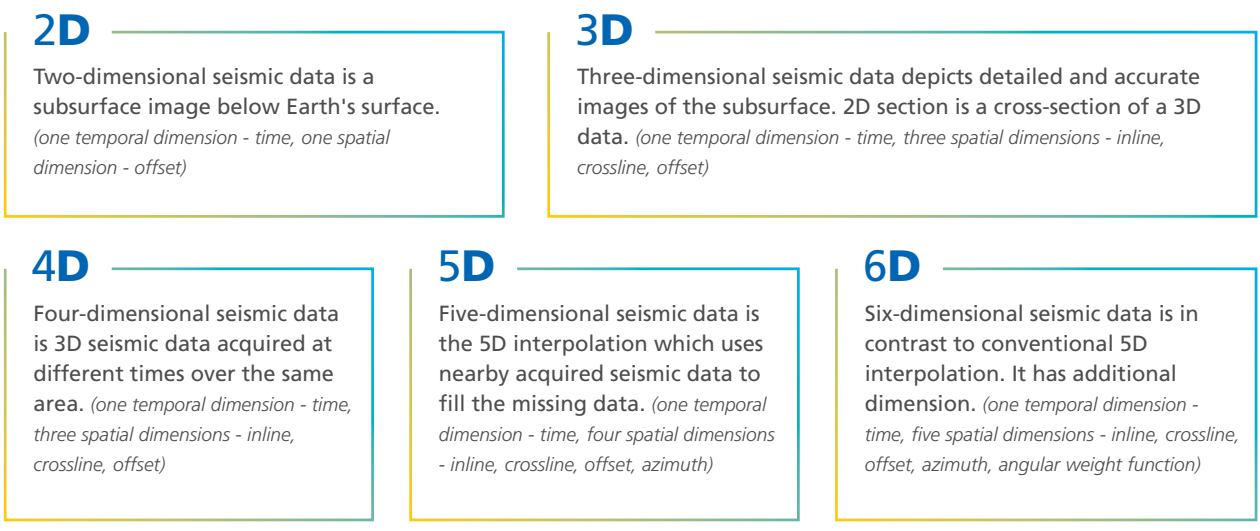


Figure 1: Classification of different seismic data types (modified after https://google.co.in)

The data can be pre-stack *(raw data gathers, intermediate processed data gathers, migrated data gathers, etc.)*, post-stack volumes of all critical processing steps, attribute volumes, and auxiliary data. It can handle high-performing workload applications with a dedicated virtual machine.

04 Migration of seismic data into OSDU through APIs

Before migrating the data into a cloud-based OSDU platform, every company in the O&G industry developing OSDU had to check the workflow process steps of ingesting seismic data into OSDU through Application Programming Interfaces (APIs). For this, one must prepare manifestations to ingest the data and follow the workflows according to OSDU standards. Seismic Store services are integrated using OSDU core services (entitlement/ obligations and storage), and Seismic Store APIs. The seismic datasets are uploaded or downloaded by using a Python utility tool called SDUTIL. This is a command line utility, one of the methods to ingest seismic data into the OSDU Data Platform. Seismic Data Management Services (DMS) uses these APIs, implemented in C++ and Python languages, to enable the ability to read and write chunks of seismic data of any size in Blob storage (Figure 2).

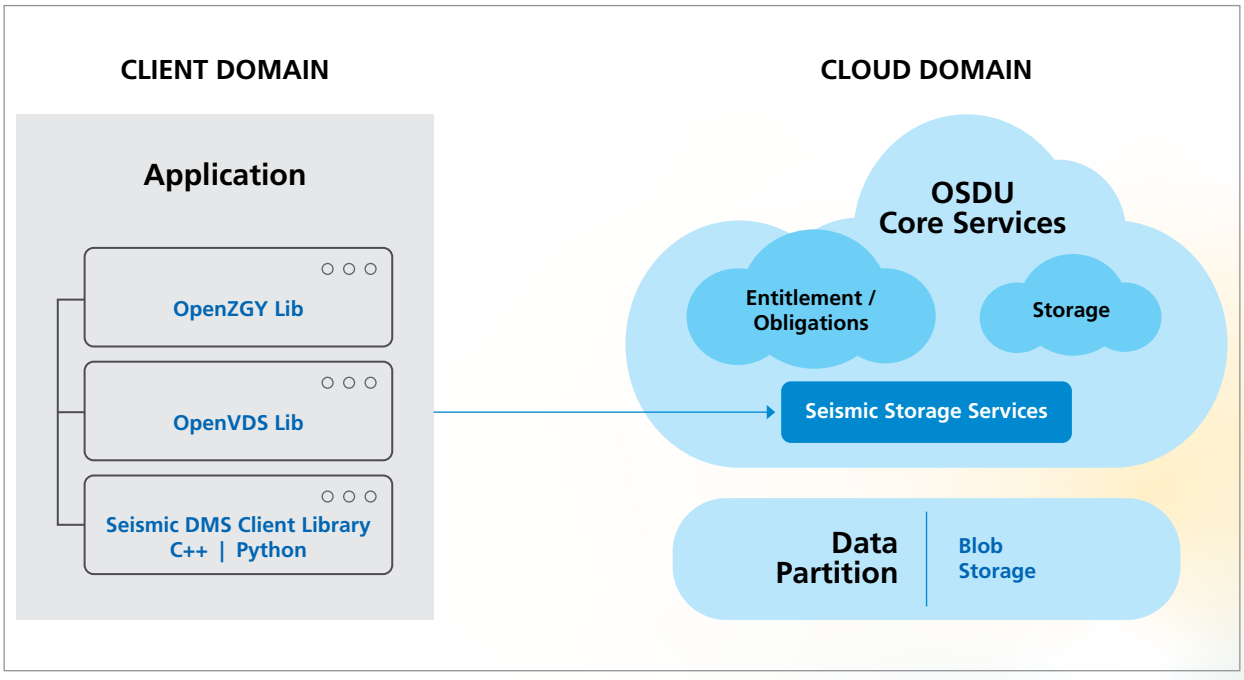


Figure 2: Seismic DMS overview (modified after https://www.youtube.com/watch?v=NlH-Eg4Yy4w&list=PL4uhUsJo0STmY4l8wRJC0Cokb9UxxVkol&index=11)

Seismic Store has three resources: Tenant (main project), Subproject (working subproject), and Dataset, which is created and registered by using SDUTIL in cloud-enabling to access the datasets securely. These three resources are addressed in the sd path as follows,

sd://<tenant>/<subproject>/<path>*/<dataset>

These could be multiple subproject resources where datasets are saved and created by Tenant admin. To access datasets in Seismic Store, users must be added or registered to a subproject resource with a specific role by the Subproject admin only. OpenVDS and OpenZGY are high-level libraries integrated with Seismic DMS client and work within the logical sd paths. Both libraries read data during the ingestion process and convert them into OpenVDS and OpenZGY formats.

Generally, all members of OSDU collaboratively work together to ingest their data through APIs acting as collections (connectors) using the Postman application, which many members of OSDU use. The following sequence describes the ingestion workflow of seismic data into the OSDU platform (Figure 3).

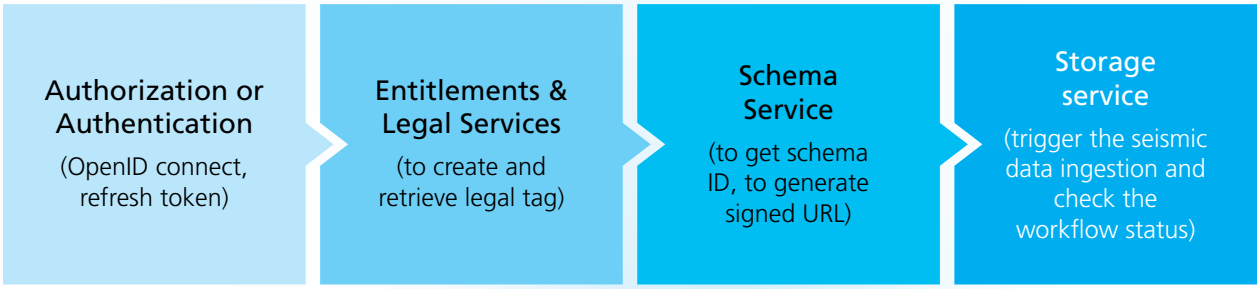


Figure 3: Steps involved in the ingestion of data into the OSDU data platform

The seismic storage service provides a set of APIs to manage the entire lifecycle of seismic data, viz., ingestion, updating, deletion, and versioning metadata. The ingested metadata is recorded as a schema containing various fields (Figure 4) and split into two parts. The first part is the basic data stored in the document database; the second part is the record payload stored in the file storage. The ingested record is indexed by the indexer service and made available via the search service.

```
{
    "id":"partition:type:uuid",
    "kind":"authority:source:entity-type:major-version.minor-version.patch-version",
    "acl":{
      "viewers":["viewers@osdu.opengroup.org"],
      "owners":["owners@osdu.opengroup.org"]
    },
    "legal":{
      "legaltags":["sample-legaltag"],
      "otherRelevantDataCountries":["FR","US"]
    },
    "data":{
      "msg":"Some Data"
    }
}
```

Figure 4: Schema defining a structure of a record across OSDU instance (modified after https://www.youtube.com/watch?v=g5oK5DxvQ30&list=PL4uhUsJo0STmY4l8wRJC0Cokb9UxxVkol&index=7)

The user provides the above record ID (record identifier), which follows the predefined naming convention while ingesting data into OSDU instance. In record ID, "kind" represents the type of data which is the name of the schema. "Access Control Level (ACL)" holds two groups of users, namely viewers and owners, who have access to this record. "legal" contains the legal constraints, such as the list of tag names and respective countries, and "data" is a record payload containing the list of key-value pairs (Figure 5).

DOCUMENT DATABASE

```
{
    "_id":"opendes:dataset--File.
Generic:7654321",
    "kind":"opendes:wks:dataset—File.
Generic:1.0.0",
    "acl":{….},
    "legal":{….},
    "tags":{},
    "status":"active",
    "user":"osdu-ingestion@osdu.
opengroup.com",
    "createTime":1678377273095,
    "modifyTime":0
}
```

FILE STORAGE

```
{
    "data":{
        "DatasetProperties":{
            "FileSourceInfo":{
                "FileSource":"xyz321",
                "PreLoadFilePAth":"https://route/
to/file",
                ….
            }
        },
        "ExtensionProperties":{….}
    },
    "meta":[{}]
}
```

Figure 5: Document database and file storage across OSDU instance (modified after https://www.youtube.com/watch?v=g5oK5DxvQ30&list=PL4uhUsJo0STmY4l8wRJC0Cokb9UxxVkoI&index=7)

The record is stored in OSDU Data Platform in two parts, i.e., document database, which contains basic data (id, kind, legal information, and access permissions), and file storage in a Java Script Object Notation (JSON) format, which contains other relevant information of the record.

For converting Seismic data (SGY or SEGY format) into Open VDS (VDS format) or Open ZGY (ZGY format) file in OSDU, one needs to configure the Apache Airflow environment variables. All the workflows related to OSDU must be licensed using Apache 2.0 License, which governs the direct code contributions. In this process, Directed Acyclic Graph (DAG), a workflow containing a set of tasks that uses a Kubernetes Pod object to run the conversion logic with containers containing an image with the DAG definition.

There are many benefits of streaming seismic data in Mercury (R3) release of OSDU once the data is stored in the cloud, namely:

- **Single source of data:** Centralized hub containing all types of bulk seismic data
- **No disk utilization:** Seismic data is streamed into applications in the local file system without being written
- **Any format:** Virtual files can be presented in different formats
- **Backward compatible:** User experience and application unchanged
- **Performance:** Streaming on VDS is much faster than data on the local system

Following a similar path, LTIMindtree also became a silver member of the OSDU forum in 2019 and collaborated with other O&G operators and OFS companies in OSDU R3 (Mercury) release. LTIMindtree, as a member of the OSDU forum, actively participates in the development of OSDU in terms of application testing, entitlements and obligations, cloud performance, especially for migrating huge seismic data in Blob storage, and integrating LTIMindtree proprietary software solutions. LTIMindtree's Exploration and Production Data Management (EPDM) platform ecosystem offers a comprehensive set of tools and services (Subsurface Digitization Workbench; GeoProQuik – MetaData Extraction, Ingestion of data; Mosaic Agnitio – Unleash the value of unstructured data; CloudEnsure – Data Platform Governance) for the deployment of OSDU Data Platform. In addition, LTIMindtree gives customers service in implementing various projects with the available software tools and encourages the customers to adopt the OSDU Data Platform.

LTIMindtree's proprietary software GeoProQuik, enables the end-user to ingest and store subsurface data in the OSDU Data Platform. The user can also use this tool's spatial query and the search feature to view and download subsurface data from the OSDU Data Platform.

05 Conclusion

In the journey of OSDU development, all O&G companies need to adapt, collaborate, and contribute their services to OSDU to benefit from each other and the end customer. As OSDU is a single centralized cloud hub with standardized VDS and ZGY formats, one can avoid the issue of standard format and storage of seismic data on the OSDU Data Platform. It can handle and store a large volume of legacy seismic data in bits and chunks by migrating into the cloud-hosted OSDU Data Platform in OpenVDS and OpenZGY format from tapes, local Disks, and Servers without any migration issues.

06 References

Moving the Subsurface into the cloud

https://osduforum.org/moving-the-subsurface-into-the-cloud/

Revolutionize Seismic Data in the Cloud

https://osduforum.org/revolutionize-seismic-data-in-the-cloud/

The OSDU™ Data Platform for everyone

https://osduforum.org/the-osdu-data-platform-for-everyone/

10 things I hate(d) about subsurface data management

https://osduforum.org/10-things-i-hated-about-subsurface-data-management/

OSDU data platform opens door to new apps in energy sector

https://www.techtarget.com/searchapparchitecture/news/252498727/OSDU-Data-Platform-opens-door-to-new-apps-in-energy-sector

R3 Ingestion workflows

https://community.opengroup.org/osdu/platform/data-flow/ingestion/home/-/wikis/R3-Ingestion-Workflows

Bluware Releases OpenVDS to The Open Group OSDU™ Data Platform

https://bluware.com/news-pr/bluware-releases-openvds-version-1-0-to-the-open-group-open-subsurface-data-universe-osdu-forum/

Understanding the Capabilities of Bluware VDS and OpenVDS

https://bluware.com/blog/understanding-the-capabilities-of-bluware-vds-and-openvds/

What is Artificial Intelligence? How does AI work and future of it

https://medium.com/@mygreatlearning/what-is-artificial-intelligence-how-does-ai-work-and-future-of-it-d6b-113fce9be

OpenVDS

https://community.opengroup.org/osdu/platform/domain-data-mgmt-services/seismic/open-vds

OpenVDS Project description

https://pypi.org/project/openvds/

Two Dimensional (2D) Seismic Survey

https://www.agl.com.au/-/media/aglmedia/documents/about-agl/how-we-source-energy/gloucester/gloucester-document-repository/fact-sheets/20110801_fact-sheet-ggp---2d-seismic-survey.pdf?la=en&hash=3F6B-698C69A86134C7217C93C11B63DE#:~:text=A%20two%20dimensional%20(2D)%20seismic,energy%20waves%20into%20the%20Earth

Introduction to 3-D seismic exploration

https://wiki.seg.org/wiki/Introduction_to_3-D_seismic_exploration

Understanding Seismic Exploration / 4D (3D-Time Lapse) Seismic Surveys

https://www.sciencedirect.com/topics/earth-and-planetary-sciences/seismic-survey#:~:text=4D%20(3D%2D-Time%20Lapse),saturation%2C%20pressure%2C%20and%20temperature.

5D Interpolation

https://www.keyseismic.com/technology/technology-catalog/5d-interpolation#:~:text=5D%20interpolation%20uses%20a%20neighbourhood,in%20the%20other%20spatial%20dimensions.

6D Interpolation of seismic data – Rationale, Practice and FAQS

https://csegrecorder.com/articles/view/6d-interpolation-of-seismic-data-rationale-practice-and-faqs#:~:text=The%206D%20interpolation%2C%20in%20contrast,in%2Ddata%2Dsupport%20situations.

Seismic DMS

https://www.youtube.com/watch?v=NlH-Eg4Yy4w&list=PL4uhUsJo0STmY4l8wRJC0Cokb9UxxVkol&index=11

Seismic store

https://community.opengroup.org/osdu/platform/domain-data-mgmt-services/seismic/seismic-dms-suite/seismic-store-service

Storage Service

https://www.youtube.com/watch?v=g5oK5DxvQ30&list=PL4uhUsJo0STmY4l8wRJC0Cokb9UxxVkol&index=6

Seismic store SDUTIL

https://community.opengroup.org/osdu/platform/domain-data-mgmt-services/seismic/seismic-dms-suite/seismic-store-sdutil

Seismic DMS C++ Client Library {SDAPI}

https://community.opengroup.org/osdu/platform/domain-data-mgmt-services/seismic/seismic-dms-suite/seismic-store-cpp-lib

Workflow DAGs installation and configuration (Open Data for industries)

https://www.ibm.com/docs/en/cloud-paks/cp-data/3.5.0?topic=data-workflow-dags-installation

Author's profile

Srinivasa Sudhakar Kadagala

Senior Specialist - Business Analysis,
LTIMindtree

Srinivasa Sudhakar is a Geophysicist with more than 15 years of experience in dealing with various methods, viz., seismic, gravity, magnetic, and electrical. He has rich experience in the acquisition and processing of seismic data. He worked with major O&G companies and is currently working with LTIMindtree as a Senior Specialist – Business Analysis. Srinivasa Sudhakar has extensive domain knowledge in acquisition and processing of seismic data, data management and is also actively involved with the OSDU testing in pre-shipping – platform validation at each milestone of a release. He has three publications to his credit in national and international journals.

LTIMindtree is a global technology consulting and digital solutions company that enables enterprises across industries to reimagine business models, accelerate innovation, and maximize growth by harnessing digital technologies. As a digital transformation partner to more than 700 clients, LTIMindtree brings extensive domain and technology expertise to help drive superior competitive differentiation, customer experiences, and business outcomes in a converging world. Powered by nearly 84,000+ talented and entrepreneurial professionals across more than 30 countries, LTIMindtree — a Larsen & Toubro Group company — combines the industry-acclaimed strengths of erstwhile Larsen and Toubro Infotech and Mindtree in solving the most complex business challenges and delivering transformation at scale. For more information, please visit *www.ltimindtree.com*.