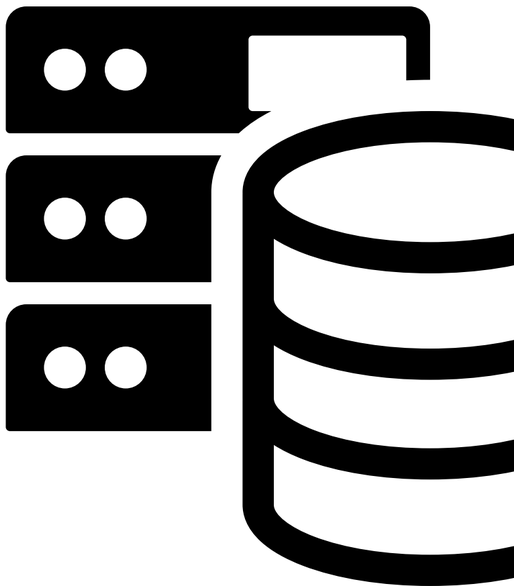
Cognizance'24 IITR

QUANTISTICO SCIENTIA

SLB DATA INNOVATION CHALLENGE

Agam Pandey |Hardik Chawla |Krish Sharma

Civil Engineering IIT Roorkee



Conceptualizing an effective metadata management strategy for *ABC* to enhance data accessibility and history tracking

Data Sources

A --> All well log files

B--> Seismic + same well log files

C--> Documents related to seismic & well log files present with A,B

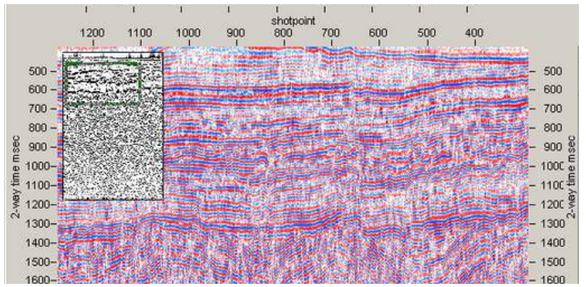
Based on metadata, ABC assumed analogous to Oil and Gas Company

- Challenges with **data management due to scattered data across multiple sources** and formats, including seismic data, well-log data, images, and documents.
- Difficult to gain an **overview** of known data, manage metadata, ensure **data accuracy**, and **security**, and promote contextual decision-making
- A **unified solution** is expected to address the above issues, emphasizing effective metadata management, enhancing data discoverability, history tracking

Oil & Gas industry- Exploration & Production

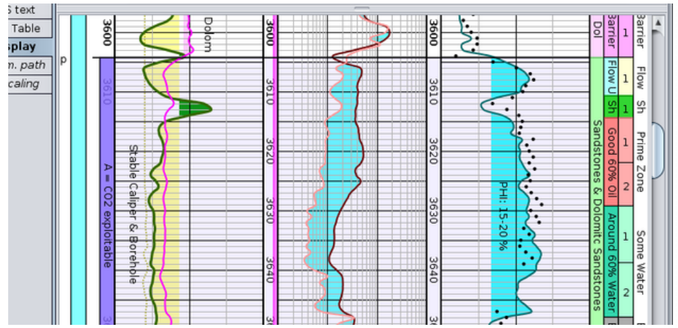
Seismic files

- Information collected through seismic surveys
- Sound waves are sent into the ground, echoes are recorded to create images (of subsurface geostructures)



Well Log files

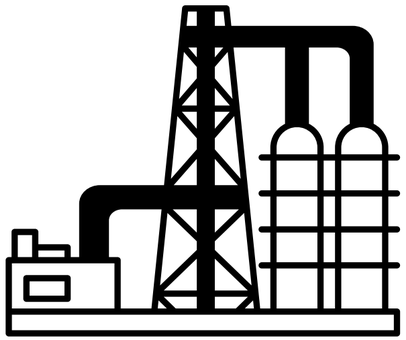
- Sound waves are sent into the ground, echoes are recorded to create images (of subsurface geostructures)



Seismic Data	.segy, .sgy
Well Log data	.las,.lis,.dlis
Images	.png,.jpeg,.tiff
Documents	.pdf,.csv,.xlsx

Understanding the supply chain of ABC with the impact of Metadata in different steps of operation

Process flow of an Oil & Gas company. Where does the metadata come from?

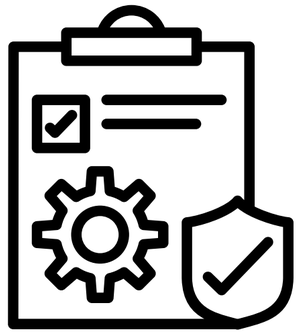


Exploration Phase

- The location,
- depth,
- quality, source, and
- context of seismic surveys,
- well logs,
- geological maps, and other data collected during exploration activities

Geological/Drilling Sites

- Well location,
- depth,
- formation characteristics,
- drilling parameters, equipment used,
- geological data gathered during the drilling process



Operational MetaData

- Data on production rates,
- fluid composition,
- well performance,
- reservoir pressure,
- equipment status,
- employee data,
- operational transactions

Distribution Phase

- Data on pipelines,
- storage facilities,
- transportation routes,
- shipment schedules, and delivery points.

Identifying underlying problems in each data flow point

- Missing data points, Incomplete records

- Large volume of data --> data being loaded becomes corrupted, inaccurate, or incomplete during the process of migration
- Ensuring data governance, security, and compliance with regulations

From multiple sources

Data Extraction → Data Quality Check → Data Migration → Data Loading → Data Warehousing

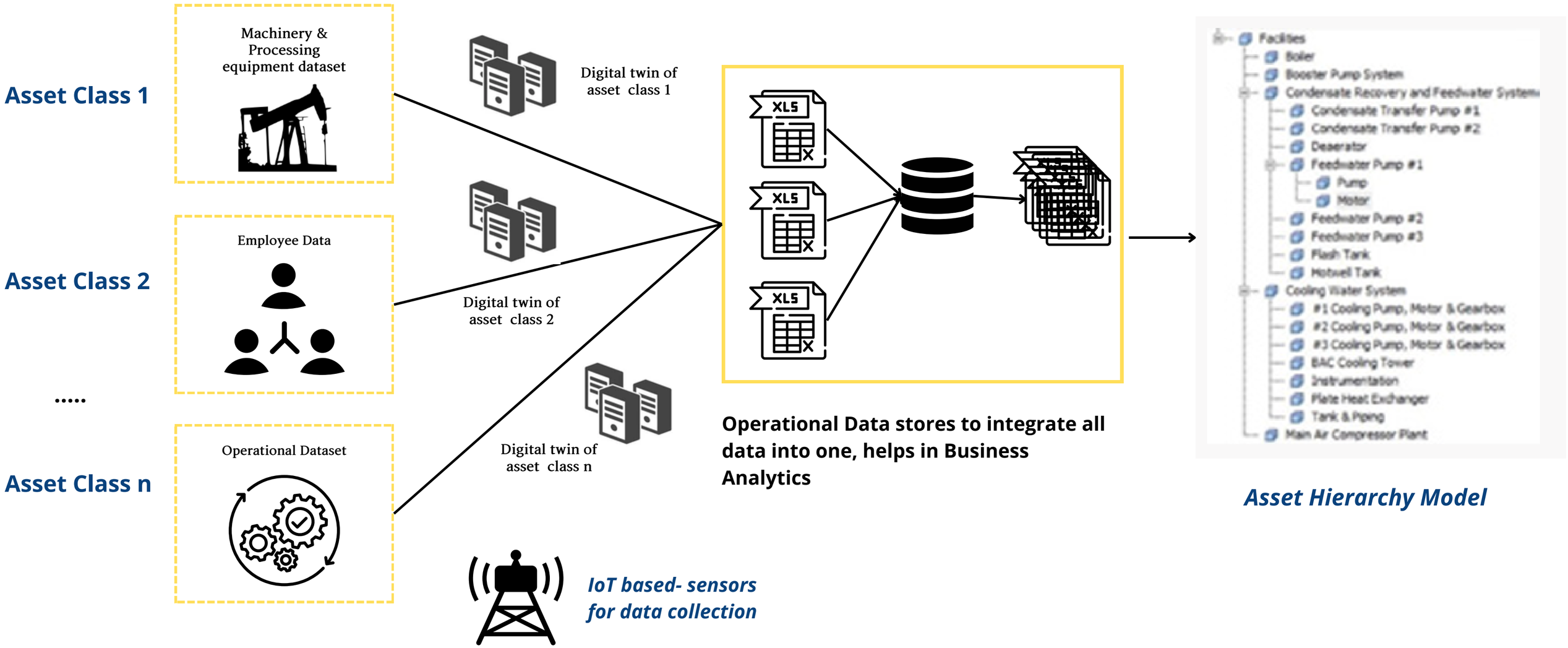
Data Analytics/ Report

- Inconsistent data formats
- Limited Automation --> Manual data entry

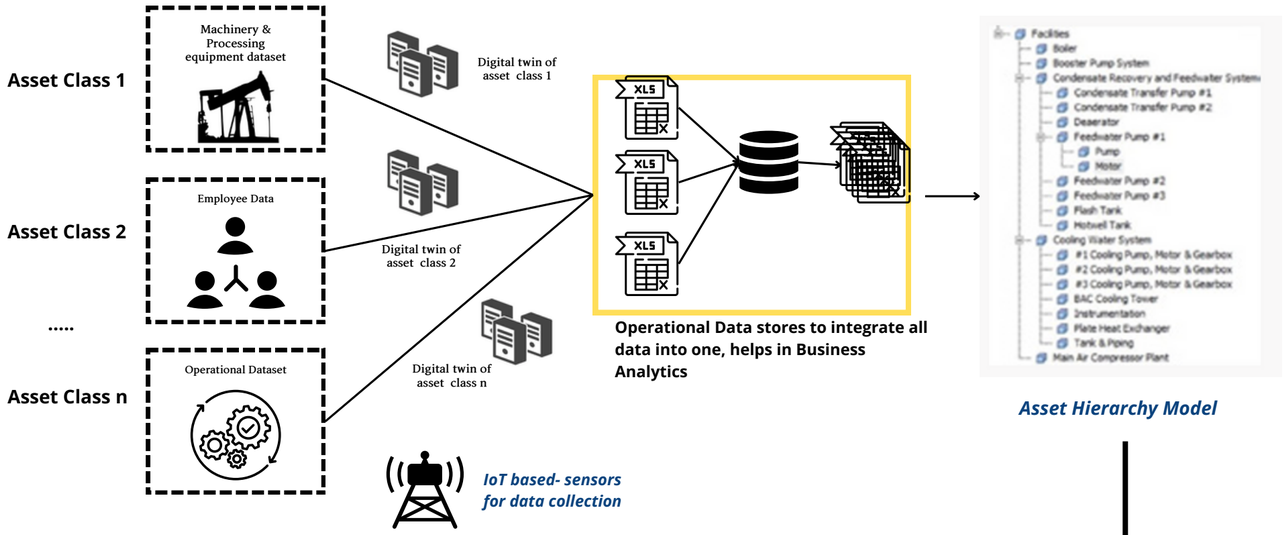
- Incompatibility between source and target systems
- Data Loss during migration

- Difficulty accessing specific data sets
- Inability to handle increasing data volumes --> Performance degradation

Operational data management---> Asset hierarchy model tackles integration, inconsistency probelms.



Operational data management---> Asset hierarchy model tackles integration, inconsistency probelms.



Asset Hierarchy Model

Operational data management

- Define the **structure and context** of sensor data, like the performance of equipment, processes, its temperature, pressure, etc. in real-time
- Enabled **analysis and integration** of all information (machinery, employee, operational) from all disparate sources
- Allow operational teams to organize information, standard KPIs, descriptive metadata and calculations

Asset Hierarchy model

- Allow users to group all information (machinery, employee, operational) in a single space
- Data consistency and accuracy of specific type of asset is ensured

Benefits of operational Data model

- Contextualized, relevant and structured information
- Avoid missing data, eliminating duplication efforts

Establishing an Centrally Version Control System

- Version control systems (VCS) for data serve a critical role in managing and tracking changes to various types of data such as seismic data, well logs, production data, engineering designs, and more

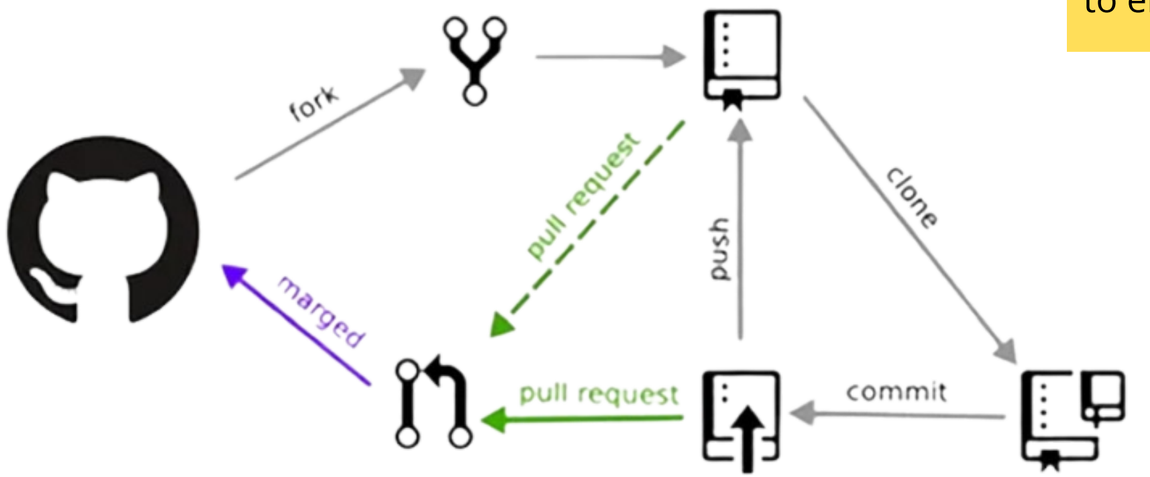
Version Control System

(A centrally managed data model)

Data Management : provides a centralized repository where data can be stored, accessed, and versioned.

Collaboration: Enable collaboration among multidisciplinary teams working on different aspects of oil and gas projects.

Audit Trail: provide a detailed history of all changes, which can be audited to ensure adherence to regulatory standards and best practices.

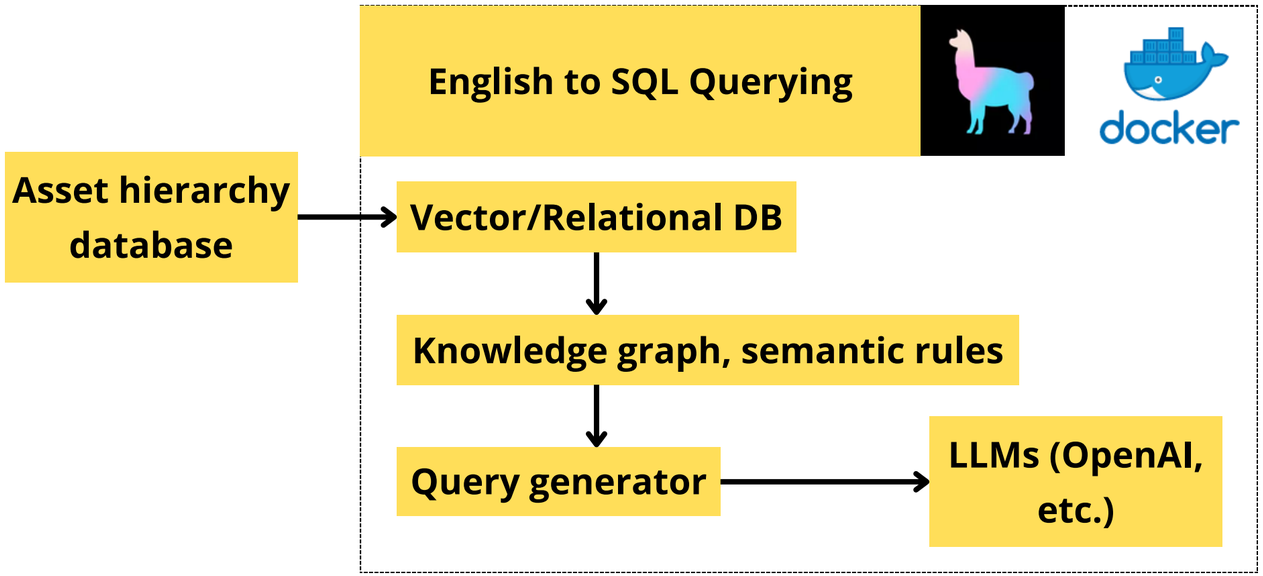


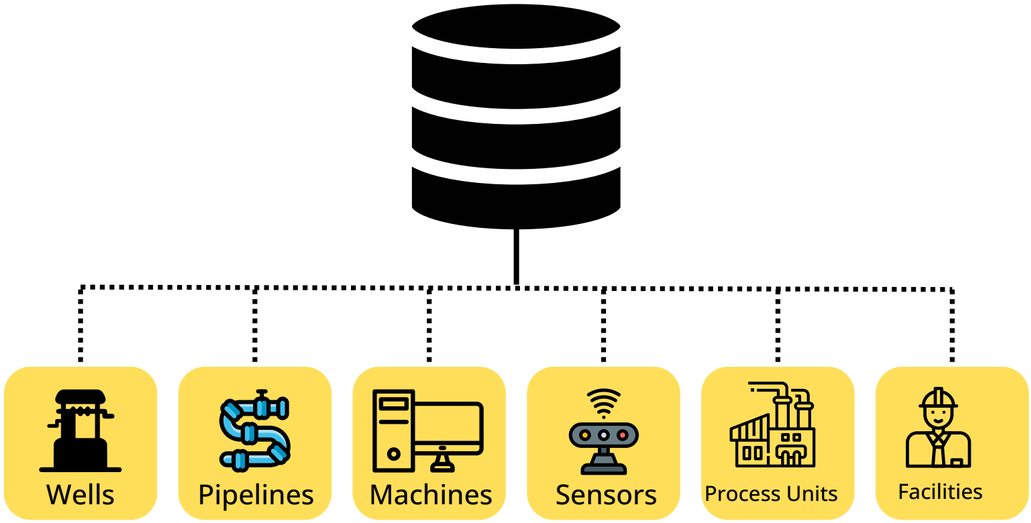
Worflow of a VCS like GitHub

Addressing the problem of data contextual interpretability, tracking history and maintenance of records over lifetime.

- Large Language Models like **Llama 70b can easily interpret structured data and unstructured dataset using LlamaIndex for text-to-SQL queries**

- Proposed solution aims at simplifying **Advanced analytics using LLMs, like LlamaIndex agent and Saas platform like waii services for storing Asset hierarchial model dataset into a vector store that can easily help in English-SQL conversation**





Advanced Analytics SaaS Platforms

dataPARC Baker Hughes

Deloitte. ENVERUS

Thank You

Agam Pandey	LinkedIn
Hardik Chawla	LinkedIn
Krish Sharma	LinkedIn