



Digital, Data Analytics, and Automation: Value Creation Through Digital E&P

19-20 NOVEMBER 2024 | BANGKOK, THAILAND

Enhancing Carbon Capture & Storage Data Management: A Novel Integration of AI, Digital Technologies, and Cross-Industry Standard Process for Data Mining Framework

The logo for AEM ENERSOL, featuring the text "AEM ENERSOL" with a red and blue graphic element.

ADAM DANIEL BIN EFFENDI
SME & PE DATA ANALYST,
AEM ENERSOL SDN BHD

&

The logo for PETRONAS, featuring a stylized teal drop shape above the text "PETRONAS".

ZAYD UMAIR ZAKARIA
PE DATA ANALYST,
UPSTREAM TECHNOLOGY, DIGITAL
& INNOVATION (UTDI), PETRONAS



1. Introduction to CCS Data Management:

- **Carbon Capture & Storage (CCS)** generates large volumes of data, but managing the associated data is complex.
- The project focuses on **centralizing Carbon Capture and Storage (CCS) data within the Storage Development Plan (SDP)**.

2. Challenges in CCS Data Management:

- **Unstructured Data:** Logs and reports are often in diverse formats, making them difficult to access, standardize, and analyze effectively.
- **Governance, Security & Scalability:** Ensuring data integrity, compliance, secure access, and scalability is crucial as CCS projects grow. Efficient systems are needed to manage increasing data volumes without compromising security or performance.

3. Project Objectives:

- **Centralize CCS Data:** Create a single source of truth for seamless data access and improved consistency.
- **Automate with AI:** Leverage low code AI to streamline data extraction and processing, reducing manual effort.
- **Governance, Security & Insights:** Ensure data compliance, scalability, and security, while enabling quick, informed decision-making using structured data.



CRISP-DM (Cross Industry Standard Process-Data Mining):

Provides a structured, repeatable approach for data mining, ensuring consistency from data understanding to deployment. It defines each step, coordinating LLM and the low code AI within the workflow.

LLM (Large Language Model):

Helps process unstructured CCS data by generating REGEX patterns, reducing manual effort during data preparation. LLM outputs are used in the low code AI to automate data extraction based on recognized patterns.



Low Code AI & Xtralisis:

1. Automates Extraction, Transformation, and Loading (ETL) processes using Xtralisis workflow (Extract files -> Detect issues ; corrupt, duplicates -> Extract/analyze text -> Classify documents -> Deliver clean, structured data for quick analysis), streamlining data extraction and ensuring efficient data readiness for analysis.
2. Low Code AI application uses the pages/data identified by LLM. ⁴

Data Accessibility

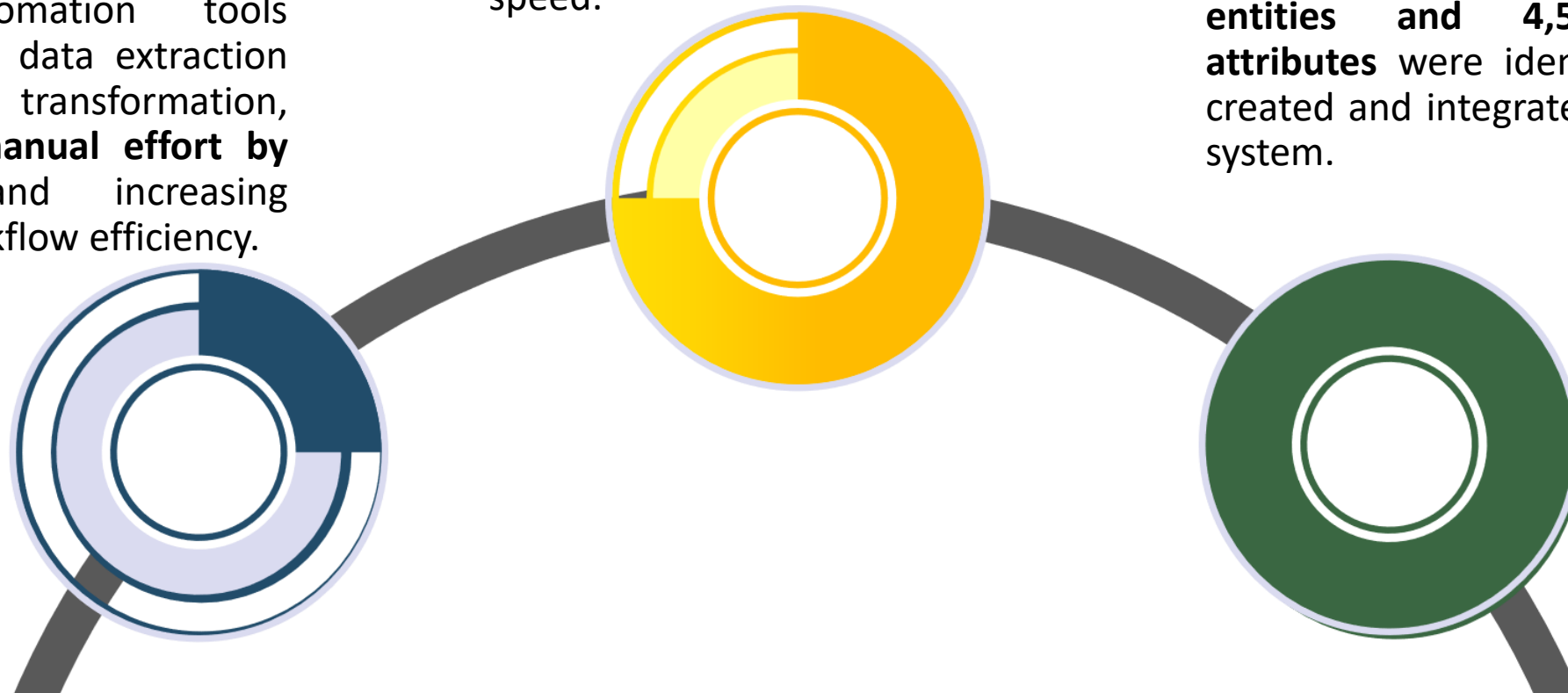
Data can now be **accessed 10x faster**, allowing project teams to retrieve structured information right away and improve decision-making speed.

Automation Efficiency

Integration of CRISP-DM with automation tools streamlined data extraction and transformation, **reducing manual effort by 10-30%** and increasing overall workflow efficiency.

Data Recovery

Approximately **261 new data entities** and **4,540 new attributes** were identified and created and integrated into the system.



Cross Industry Standard Process for Data Mining (CRISP-DM)

1. Business Understanding:

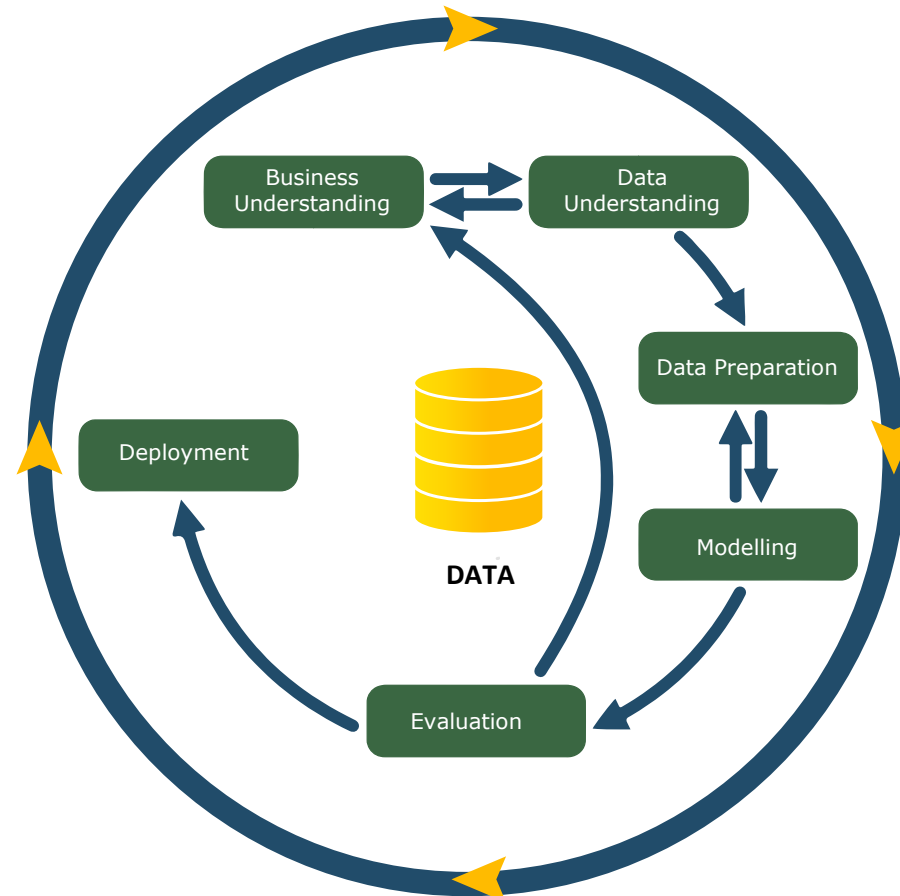
- **Project Goal:** Modernize CCS data management.
- **Alignment:** Focus on creating a centralized database.

2. Data Understanding:

- **Data Challenges:** The project faced **unstructured data**.
- **Assessment:** Identified key data points needed for analysis.

3. Data Preparation:

- **Automation:** Low code AI was used to **automate data extraction and transformation**.
- **Transformation:** Using **LLM-assisted REGEX**



4. Modeling:

- **Reduction of Manual Efforts:** Automation through low code AI.

5. Evaluation :

- **Model Evaluation:** Ensured that the automated data extraction met the quality.
- **Assessment:** Identified key data points needed for analysis.

6. Deployment:

- **System Implementation:** The processed and structured data was ingested into **CCS Storage Development Plan (SDP) database**.

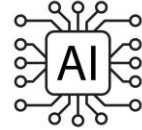
Large Language Model (LLM)



- 1. LLM for REGEX Creation:** LLMs helped generate and validate complex REGEX for more flexible and precise data extraction.
- 2. Key Term Identification:** LLM helped create REGEX tailored to our understanding of the document, allowing efficient data location within large reports
- 3. Data Security:** A locally deployed LLM ensured data privacy, reducing reliance on cloud-based services.

ATTRIBUTE	EXAMPLE VALUE	SONNET 3.5 CLAUD REGEX	GPT-4 REGEX	LLM INTERPRETATION DIFFERENCE
Well Name	ABC-1	<code>^[A-Z0-9-]+\$</code>	<code>([A-Z]+\-\d+)</code>	SONNET: Fixed pattern. GPT-4: Capturing groups for naming conventions.
Bottom Logged Interval	1500.29	<code>^\d+\.\d{2}\$</code>	<code>(\d+\.\d{2,})</code>	SONNET: Restricts to two decimals. GPT-4: Flexible with decimal places.
Log Description	MUD LOG TVD-MSL	<code>^[A-Z\s-]+\$</code>	<code>([A-Z\s]+LOG\s[TVD-MSL]+)</code>	SONNET: Simple pattern. GPT-4: Context-aware for expected structures.
Scanned QC Remarks	ABC-1-1_MUDLOG 44.90-1500.29 TVD-MSL	<code>^[A-Z0-9_\s-]+\$</code>	<code>([A-Z0-9_\s-]+_MUDLOG\s\d+\.\d+\d+\.\d+\sTVD\MSL)</code>	SONNET: Basic remarks. GPT-4: Captures complex fields, numeric ranges, and abbreviations.

Low Code AI



Define the Project and Data Needs

- Select the fields or attributes to be extracted (e.g., well name, depth, production dates).
- Define the format for the final CSV output.

Train Low Code AI with Document Variations

- Document Collection
- Machine Learning in the application
- Training Iteration



Automated Data Extraction

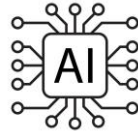
- The application uses the trained ML model to identify and extract key data fields.
- The application performs basic validation during extraction, checking that key fields are filled and formatted correctly

Data Storage

- The extracted data is automatically populated into a CSV file and/or database.



Low Code AI



Quick test

for 8.5" section production hole. In addition, Halliburton Geosteering crew comprising two engineers were hired when during the drilling of EB-A115T1 (Figure 6). They were based in the Kuala Lumpur SPE office. The production hole (EB-A115T1) was sidetracked from 698.9m and drilled to final TD at 2204m MD/1164.6m TVD (-1119.6m TVDSS) on 27 April 2016 at 11:45 hrs with total length of horizontal section 1658m (total at 1546m MD).

PROGNOSIS		PRE-DRILL		ACTUAL		DIFF (m-TVD)		Remarks
m MD	m TVD	m MD	m TVDSS	m MD	m TVD	m MD	m TVD	
9-5/8" Sidetrack Window	709	892	626.0	688.9	679.2	624.0	-57.2	
1-80 Coal	805.6	789.2	744.0	811.5	785.5	741.3	-44.2	
1-10	890.9	814.7	789.5	727.0	820.9	789.8	-31.1	
1-18	1016.7	933.2	888.0	1012.0	931.0	889.8	-22.2	
1-18	1150.5	1022.2	977.3	1143.0	1016.3	971.1	-34.4	
6-5	1241.5	1092.5	1050.0	1232.3	1077.0	1031.8	-32.2	
6-8	1316.6	1189.2	1074.0	1309.0	1136.7	1071.5	-25.5	
8-10 Upper Delta	1442.0	1197.2	1117.0	1410.0	1193.5	1189.3	-3.7	None
Gas Oil Contact (GOC)	1465.0	1162.2	1117.0					
8-10 Lower (Coastal Plate)	1465.0	1165.2	1120.0	1462.0	1164.2	1119.0	-1.0	
Landrig point	1523.8	1165.2	1120.0	1548.0	1164.8	1119.6	-0.4	
Final TD	2000.0	1164.7	1119.5	2024.0	1164.8	1119.6	0.1	

Learn how you can improve your model [Start over](#) [Close](#)

Quick test

PROGNOSIS	Reservoir name	TD MB/AB	TVDDF	TVDSS	Remarks
9-5/8" Sidetrack Window	9-5/8" Sidetrack Window	698.9	679.2	634.0	
1-80 Coal	1-80 Coal	821.5	785.5	741.3	
1-10	1-10	877.0	822.0	786.6	
1-18	1-18	1012.0	951.0	895.8	
6-5	6-5	1143.0	1016.3	971.1	
6-8	6-8	1232.3	1077.0	1031.8	
8-10 Upper Delta	8-10 Upper Delta	1309.0	1136.7	1071.5	
Gas Oil Contact (GOC)	Gas Oil Contact (GOC)	1410.0	1133.5	1104.3	Not observed
8-10 Lower (Coastal Plate)	8-10 Lower (Coastal Plate)	1462.0	1164.2	1119.0	
Landrig point	Landrig point	1548.0	1164.8	1119.6	
Final TD	Final TD	2004.0	1164.8	1119.6	
8-10 Upper Delta	8-10 Upper Delta	1410.0	1133.5	1104.3	

Fluid contact:
No pilot hole was created in this well. Current Gas Oil Contact (GOC) at 8-10 Sandstone was not observed in the well due to poor rock quality.

East Belumat-A115T1 - Reservoir Summary (RT Data)

Learn how you can improve your model [Start over](#) [Close](#)

Tagging the information in the unstructured report in **AI Builder**

Observed and QC the extracted information/data.

Export the information into **CSV** for further transformation and QC

The collaboration between PETRONAS and AEM Energy Solutions led to the successful automation of CCS Data Extraction and Transformation using the CRISP-DM framework.

- **Improved Data Accessibility:** Enabled 10x faster data retrieval, allowing project teams to access structured information more efficiently, supporting quicker decision-making.
- **Automation Efficiency:** CRISP-DM integration streamlined data extraction, reducing manual effort by utilizing Low Code AI, and increasing workflow efficiency.
- **Enhanced Data Quality & Recovery:** Approximately 261 new data entities and 4,540 attributes were accurately identified and integrated, ensuring reliable datasets for analysis.

The system is scalable and designed to support future CCS projects effectively.

"Empowering Data, Driving Efficiency—Securing the Future of CCS."