Genai Accelerated Upstream Data Platform, From Onpremise Ingestion To Cloud Insights

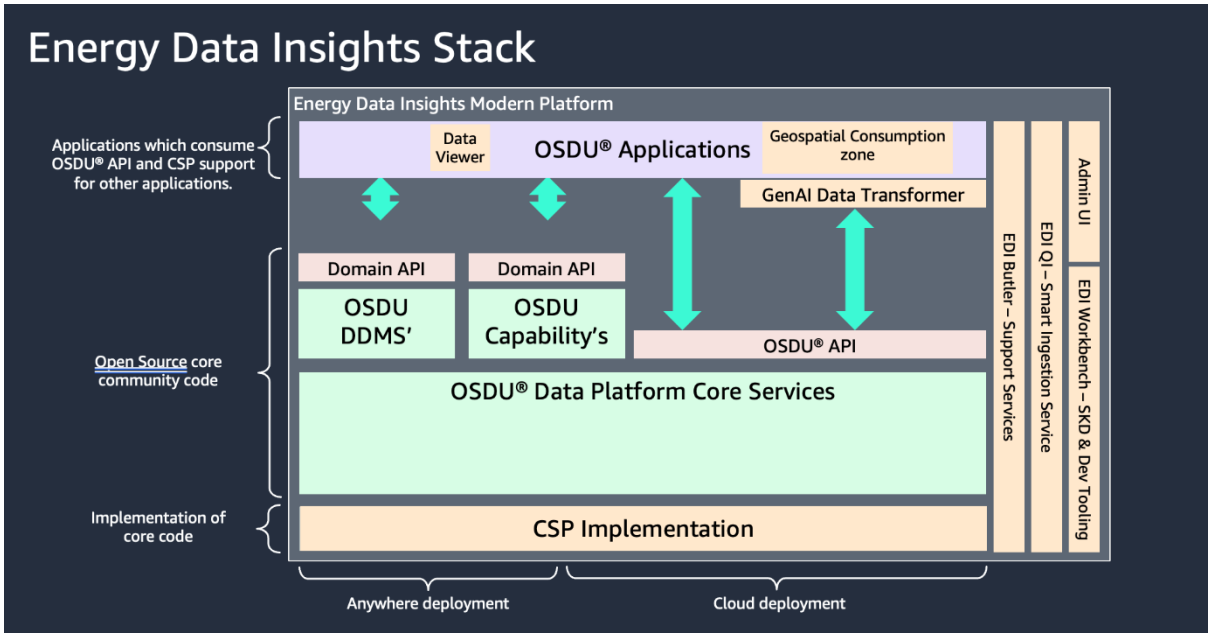
W.A. Shaari, Amazon Web Services

Objectives/Scope: A framework for rapidly ingesting and harmonizing upstream data from on-premises sources into cloud-native platforms using GenAI-powered automation. The scope covers schema mapping, metadata enrichment, and real-time data integration, enabling seamless analytics and decision-making for exploration and production teams while ensuring data residency and sovereignty compliance.

Methods, Procedures, Process: The approach leverages a GenAI-enabled ingestion module that automates mapping, transformation, and validation of subsurface data from legacy repositories to OSDU-compliant cloud environments. The process uses parallelized computing, automated lineage tracking, and schema harmonization to accelerate onboarding. User experience is enhanced by GenAI-driven interfaces for data discovery and workflow automation. The methodology is validated through benchmarking ingestion speed, data quality, and domain application integration, considering regulatory and operational constraints in O&G environments.

Results, Observations, Conclusions: Results show the GenAI-powered ingestion framework reduces onboarding timelines from months to weeks, achieving 40% improvement in data mapping accuracy and enhanced data discoverability. Automated schema harmonization enables cross-domain analytics and rapid integration with upstream applications, while metadata lineage ensures traceability and compliance. GenAI-driven interfaces significantly lower barriers for both technical and non-technical users to access complex subsurface datasets. The approach facilitates real-time analytics, supporting faster decision cycles and improved collaboration across exploration, engineering, and business units. Conclusions demonstrate that integrating GenAI within data ingestion breaks down legacy silos while establishing a scalable foundation for enterprise-wide digital transformation in upstream oil and gas.

Novel/Additive Information: A GenAI-driven ingestion and integration workflow purpose-built for upstream O&G, moving beyond traditional ETL by enabling real-time, schema-aware, and user-centric data onboarding. Engineers gain actionable strategies to bridge on-premises and cloud data, streamline integration, and accelerate analytics adoption while maintaining compliance in regulated, data-sensitive


Energy Data Insights Stack
Energy Data Insights Modern Platform
Applications which consume OSDU® API and CSP support for other applications.
Data Viewer
OSDU® Applications
Geospatial Consumption zone
GenAI Data Transformer
Domain API
Domain API
OSDU DDMS'
OSDU Capability's
OSDU® API
Open Source core community code
OSDU® Data Platform Core Services
Implementation of core code
CSP Implementation
EDI Butler – Support Services
EDI QI – Smart Ingestion Service
Admin UI
EDI Workbench – SKD & Dev Tooling
Anywhere deployment
Cloud deployment