Data Ingestion Pipelines for Seismic Crew

Mohit Lokwani, Vijaykumar Raj
QC - IT, Arabian Geophysical and Surveying Co. (ARGAS), King Fahad Road, 34232, Dhahran, Saudi Arabia
vraj@argas.com

Introduction

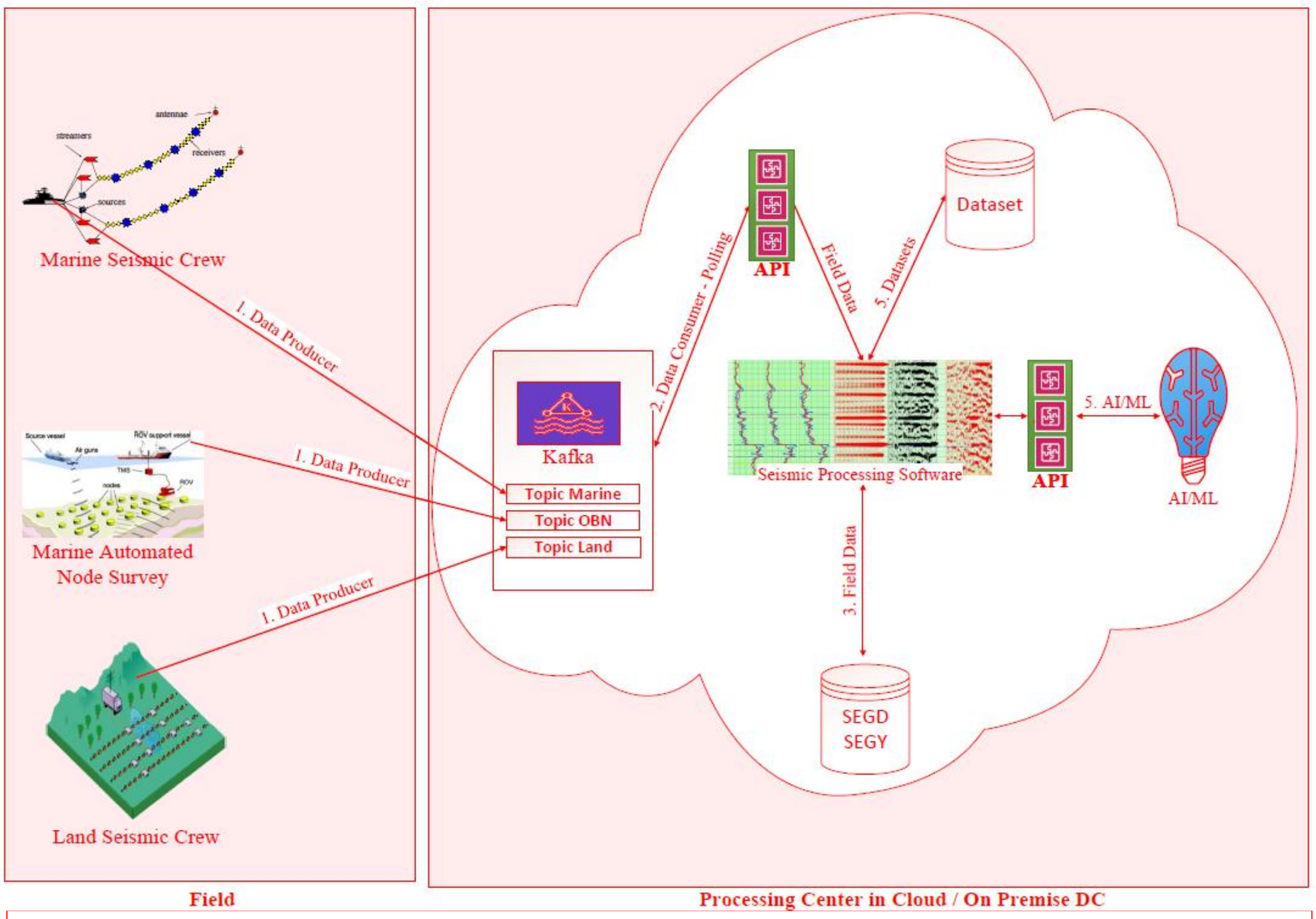
This research explores integrating real-time streaming platforms like Kafka with processing software via APIs. The processing software, leveraging AI and machine learning, automates the processing of seismic data collected by the seismic crew for oil and gas exploration. Vibrations created by energy sources like dynamite or vibrator, create waves that are directed below the surface. These waves get reflected or refracted depending on the geology of that place and are captured by geophones(onshore) or hydrophones(offshore). Acquired data is subjected to regular quality checks so that any anomalies are corrected through reshooting if required. The processed data with an improved signal-to-noise ratio provides a clear subsurface picture for better interpretation of oil and gas reserves.

Objective

This research aims to automate the processing of seismic data by connecting real-time data streams(Kafka) with processing tools via API integration. The acquired seismic data is continuously filtered and corrected using Machine Learning and AI so as to improve signal-to-noise ratio. Henceforth, automating Quality Control(QC) and processing sequences on acquired data to produce sharper subsurface images. Ultimately, the goal is to streamline data collection and automate the processing of seismic data.

Method Procedures and Process

The research integrates Kafka into seismic data acquisition via API-enabled processing software. Data collection involves monitoring vibrations and converting waves to digitized data. Quality checks with the help of filters and corrective measures are implemented, with Kafka enhancing efficiency. The signal-to-noise ratio is optimized by the processing software by utilizing Convolutional Neural Networks(CNNs) and Wavelet-based approaches. This ensures a clear interpretation of subsurface characteristics for a better understanding of oil and gas reserves.



1. Data Ingestion with Kafka:
- Seismic data from field and marine sources is captured by a recorder, acting as a Kafka producer.
- The recording application publishes this data to specific Kafka topics, with each crew assigned a separate topic.
- Topics are divided into multiple partitions to enhance parallel processing and scalability.

2. Real-Time Data Access:
- A processing application, acting as a Kafka consumer, accesses real-time field data via APIs as soon as new data is ingested into Kafka.
- This enables immediate processing and reduces latency in data availability.

3. Data Storage and Direct Processing:
- Ingested data is stored for further processing; however, direct processing from Kafka topics is also possible, enabling the generation of datasets on-the-fly.

4. AI/ML Module Integration:
- The processing application dynamically loads necessary AI/ML/DL modules via APIs based on the field data received.
- This modular approach ensures the application is equipped with the right tools for each processing task.

5. Advanced Seismic Processing:
- AI/ML/DL modules are utilized to enhance seismic data processing, focusing on improving SNR, conducting velocity and attribute analysis, detecting faults, performing seismic inversion, and interpreting results.

Results Observation and Conclusion

The integration of Kafka with processing software via API transforms seismic data processing. Continuous data flow, rapid anomaly detection, and reduced reshooting increase its efficiency. Artificial intelligence and Machine Learning elevate signal-to-noise ratios, yielding accurate subsurface images. This seamless integration streamlines processes, conclusively enhancing efficiency, accuracy, and cost-effectiveness in seismic data acquisition for oil and gas reserves.

Additive information

This research pioneers a transformative synergy between Kafka, AI-powered processing, and CI/CD pipelines. With dynamic filters and corrections, it automates and enhances the signal-to noise ratio, while optimizing oil and gas reserve interpretation.

Apache Kafka is a distributed event streaming platform used for building real-time data pipelines and streaming applications. It is designed to handle large volumes of data with low latency, making it ideal for applications that require high-throughput and fault-tolerant messaging. Core component of Kafka are
Producer, Consumer, Broker, Topic, Partition, Zookeeper

Different AI ML DL models that can be utilized for data processing
Noise Reduction – Autoencoders, Wavelet Transform, Generative Adversarial Networks (GANs)
Signal Enhancement, Migration, Fault Detection and Intepretation – Convolutional Neural Networks (CNNs)
Velocity Analysis – Linear Regression, Polynomial Regression, Support Vector Regression (SVR)