Harnessing the Power of a Fully Managed Universal Data Lakehouse:

Solving Complex Data Challenges
for Predictive Maintenance at Olameter

ONEHOUSE × *Olameter*

Contents

Advancing Predictive Maintenance in Utilities

Predictive maintenance uses data and advanced analytics to anticipate and prevent equipment failures. For Olameter, a provider of communication and network monitoring for electric, water, gas, and telecommunication utilities, as well as energy retailers and management entities, the initial step to develop predictive maintenance systems is to establish the topology of the energy infrastructure.

This involves collecting data from individual electricity meters and existing transformers, and then determining which meters are connected to which transformers. Through telemetry data analysis and correlation, Olameter can accurately map the network setup and use predictive models to forecast potential power outages.

Olameter

Olameter is a leading provider of outsourced utility asset management, network communications, and data services. With a staff of more than 1,500 technical, administrative, and field service professionals, Olameter serves over 350 clients across North America. Olameter provides communications and network monitoring needs for the electric, water, gas and telecommunication utilities, as well as energy retailers and energy management entities. The company's specialty is generating cost savings and operational efficiencies for their clients by integrating services, technology, and personnel across multiple clients.

Olameter's Challenge: Tackling Massive Data Processing

The Roadblock: Processing Vast XML Data Efficiently

However, on the path to predictive maintenance, Olameter hit a major roadblock: processing XML data packed with valuable meter reading information. They initially tried using custom .NET applications to parse and restructure these XML files, but it quickly became clear this wouldn't work. The massive amount of data would take over six months of compute time just to process a single year's worth of information, making the needed analysis both financially and logistically impossible.

Yet, this data was vital for predicting outages, enabling preventive measures, and providing clear network topology visualizations. The bottleneck caused nearly a decade's worth of XML data to build up, unprocessed. Faced with this challenge, Olameter teamed up with Onehouse. Together, they embarked on a journey to bring the XML data into the Onehouse Universal Lakehouse. The results were game-changing.

Solution: From XML Ingestion & Transforming Data Layouts to Advanced Data Pipelines

Custom XML Ingestion: Achieving Record Processing Times

Within two weeks, the Onehouse team developed a custom XML ingestion solution for Olameter, which is now available to all Onehouse users. This solution allowed Olameter to intelligently parse through XML datasets - breaking up relevant struct and array objects to land in analytics-ready lakehouse tables. This breakthrough allows Olameter to break open their complex XML files and start running SQL queries on the data, all via a no-code/low-code interface.

The interface to this custom XML ingestion solution can be seen in the figure below.
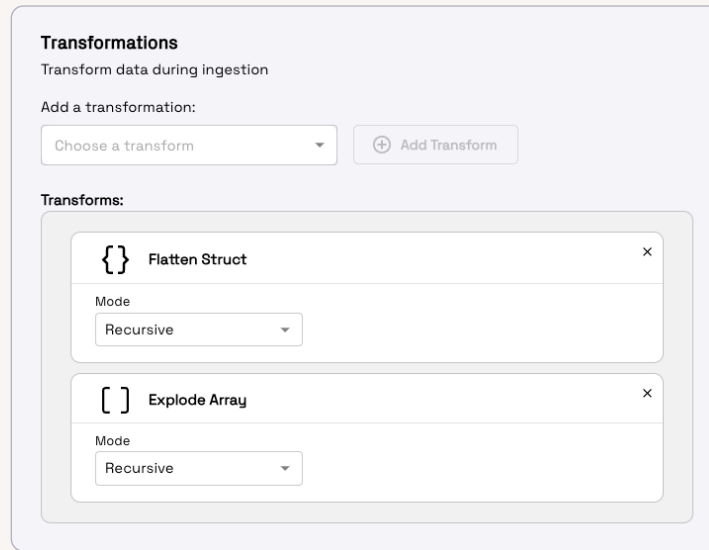

Figure 1: Onehouse complex object parsing

Leveraging Apache Hudi under the hood, Onehouse enables Olameter to significantly enhance their XML data management capabilities. Onehouse's incremental data processing framework allows Olameter to handle their nearly decade-long data backlog in manageable chunks, and handle new data as it arrives, transforming a computationally intensive challenge into a feasible, long-term solution.

Onehouse's built-in support for upserts and deletes ensures that updates to XML data and other enrichment datasets are seamlessly reflected in Olameter's lakehouse tables, eliminating the need for full dataset rewrites. Additionally, the storage optimization capabilities of Onehouse enable efficient management of large-scale data across multiple versions without the overhead of complex lifecycle policies.

Onehouse's support for Parquet and ORC file formats facilitates fast query performance and efficient, compressed storage. Seamless integration with the Hadoop ecosystem, including tools such as Apache Spark, further streamlines the processing and querying of Olameter's XML data.

This architecture significantly enhances Olameter's data processing capabilities, making the XML data easily accessible and actionable, and transforming a previously daunting task into an efficient and manageable process.

Flattening and Exploding: Optimizing Data Formats

Flattening deeply nested objects involves transforming hierarchical or nested data structures, like those in JSON or XML formats, into a simpler, tabular format. This iterative process employed different optimization layers to minimize both data storage size and query time. The first layer was tied to the business need and the XML data structure, focusing on flattening only relevant data for analysis while encapsulating non-relevant data into arrays. This was done by analyzing XML fields, with reference to the documentation provided by the data collection system provider.

The second optimization layer involved partitioning the data to query only relevant portions. Since electric meters and transformers are geotagged, and given the typical distances between each transformer and its connected meters, Olameter used geo-spatial clustering to partition the meter data. This process makes the data easier to query and analyze by bringing deeply nested elements to the top level, resulting in a flat geo-spatial partitioned table where each row represents a single entity and each column represents an attribute.

Original Nested Structure

id	name	address
1	John Doe	{"street":"123 Main St", "city":"Anytown", "state":"CA", "postal_code":"12345"}

Flattened Table Structure

id	name	address_street	address_city	address_state	address_postal_code
1	John Doe	123 Main St	Anytown	CA	12345

Figure 2: The process of flattening

Exploding large arrays involves taking elements within arrays or lists and expanding them into individual rows in a table. For example, exploding an array for a record containing a list of items creates separate rows for each item, allowing each element of the array to be treated as an individual record. This facilitates more granular analysis and querying.

Through these transformations, Olameter created a clear pathway to extract value from a decade's worth of XML meter data. They achieved their desired data model in just six weeks, significantly enhancing data processing and querying efficiency.

Identifying and Solving Processing Inefficiencies: Optimizing Tables and Processes

The collaboration didn't stop there. Olameter worked closely with the Onehouse team to scale the data model and identify processing inefficiencies. Analyzing compute and query usage patterns, Onehouse recommended XML parsing solutions that would result in only the needed data structures being processed, while data not currently in use remained nested.

The result? A 10x saving on compute. Olameter has now implemented a Medallion data architecture and a range of best practices in data engineering.

Building Downstream Pipelines and ML Models

Olameter enjoys substantial benefits from its deep relationship with Onehouse, which prides itself on forming robust design partnerships with its customers. Through this collaboration, Onehouse advises Olameter's data science and engineering teams on building downstream pipelines and ML models, leading to the creation of silver and gold tables that greatly enhance Olameter's data processing and querying capabilities.

Moreover, Olameter benefits from Onehouse's 24/7 on-call engineering support, ensuring that any broken pipelines trigger immediate technical collaboration. For compelling use cases, Onehouse engineers work hand-in-hand with Olameter's teams to develop innovative features that unlock significant business value, such as new ingestion protocols, databases for CDC, low-code transformations, and advanced downstream analytics and ML solutions.

Results: Accelerated Data Processing from Years to Days with Near Real-Time XML Ingestion and Seamless Scaling

The partnership with Onehouse transformed Olameter's XML challenging use case into a cost-efficient and robust business solution. Data processing times were reduced from months to days, and incremental XML data can now be processed in near real-time with seamless elastic scaling and minimal infrastructure management. Olameter has gained robust processing pipelines enriched with machine learning models for fast querying and operational insights into outages and meter telemetry, turning their data challenges into significant business advantages.

"The ability to process incremental XML data in near real-time has been a game-changer. With Onehouse, we can now leverage machine learning models to gain rapid insights into outages and meter telemetry, enhancing our operational efficiency."

Ph.D. Taieb Lamine Ben Cheikh,
Data scientist, Olameter Inc.

Summary of outcomes:

Managed XML File Ingestion:	A first in the industry.
Processing Time Reduction:	From months and years to hours and days.
Custom Optimizations:	Cut data processing times by a further 10x.
Robust Pipelines:	Co-designed by Onehouse for XML analytics.
ML-Based Geo-Clustering:	Push Onehouse data to downstream models.
Operational Insights:	Achieved through advanced data processing and querying capabilities.

Through this collaboration, Olameter and Onehouse are redefining what's possible in data processing and analytics. They set a new standard for efficiency and innovation against a challenging use case, proving the power of a well-executed partnership and the transformative potential of a fully-managed Universal Data Lakehouse architecture.

Experience the Universal Data Lakehouse Advantage:

onehouse.ai/schedule-a-test-drive