Claudia V Dominguez - Notes
Oil Field Monitoring System Notes

Synthetic dataset modeled after typical PDVSA production assets (pump jacks, ESPs, compressors). No proprietary or real operational data used

- Sites grouped by basin: Maracaibo (Zulia) vs Orinoco Belt (Anzoátegui/Monagas)

- Work orders / alerts labeled like: "Maintenance Ticket", "Crew Dispatch Priority"

- Dashboard sections: Field Ops, Integrity, Reliability, Production

- Local time view: VET (UTC−4) plus UTC stored in DB

System Architecture

End-to-end pipeline:

1. Data Simulation and Ingestion (Python)

    - Generates synthetic sensor readings for oil-field equipment

    - Ingests data at regular intervals (acceptable 1–5 minutes)

2. Cloud Data Storage (Azure SQL Database)

    - Stores raw sensor telemetry

    - Stores detected anomalies and health metrics

3. Analytics and Visualization (Power BI)

    - Provides near real-time dashboards

  - Enables drill-down analysis per asset and site

Data Model Overview

Core Tables

- dim_asset
  Stores metadata about oil-field equipment (site, basin, equipment type).

- dim_sensor
  Defines sensor types and measurement units.

- fact_sensor_reading
  Time-series sensor telemetry for each asset.

- fact_anomaly_event
  Records detected anomalies with severity and scores.

- asset_health_score
  Aggregated risk score indicating likelihood of failure.

<u>Design Principles</u>

- Star-schema inspired modeling

- Time-series optimized indexing

- Separation of raw vs analytical data

<u>**Architecture style:**</u>
Batch ingestion with frequent refresh (near real-time).

Sensor data is simulated to reflect realistic oil-field behavior:

- Normal operating ranges per sensor

- Gradual drift patterns
- Sudden spikes and flatlines
- Occasional missing readings

Ingestion flow:

1. Generate or load sensor readings
2. Validate schema and data types
3. Insert readings into Azure SQL
4. Trigger anomaly detection logic

All timestamps are stored in UTC, with local Venezuelan time (VET) derived during visualization.

Anomalies are detected using explainable statistical methods:

- Rolling Z-score analysis
- Rate-of-change thresholds
- Flatline detection (sensor failure)
- Missing data detection

Each anomaly is classified by:

- Type (spike, drift, flatline, missing)
- Severity (Low / Medium / High)
- Anomaly score

To estimate potential failure risk:

- Anomalies contribute weighted points based on severity
- Scores decay over time when no new anomalies occur
- Assets exceeding a threshold are flagged as high risk

Dashboard Pages

1. Fleet Overview

- Total assets monitored
- Active anomalies (last 24 hours)
- High-risk equipment list

2. Asset Drilldown

- Sensor trends over time
- Anomaly markers on charts
- Asset-level health score

3. Maintenance Insights

- Anomaly distribution by site and basin
- High-priority maintenance candidates

Refresh Strategy

- Scheduled refresh every 15–30 minutes
- Supports near real-time monitoring scenarios

<u>Possible Enhancements</u>

- Streaming ingestion with Azure Event Hubs
- Machine learning-based failure prediction
- Alerting via email or SMS
- Historical downtime and cost impact modeling

Disclaimer: This is a portfolio case study using simulated sensor data. It is not affiliated with PDVSA, and it does not use any proprietary or confidential information.

Sources:
https://llumin.com/blog/predictive-maintenance-oil-and-gas

https://innovamas.nakasawaresources.com/en/optimizing-and-monitoring-a-heavy-oil-reservoir-using-progressive-cavities-pumps-and-real-time-information-in-dobokubi-field-faja-petrolifera-del-orinoco-ayacucho-division-%E2%80%90-venezuela