



Data Transfer for Oil & Gas Operations POV Case Study

The energy transition narrative has shifted. In late 2025, the International Energy Agency reversed its long-standing prediction that oil demand would peak before 2030. Under its current projections, global oil demand will reach 113 million barrels per day by 2050 — a 13% increase from 2024 levels. Rising energy consumption from AI data centers, petrochemicals, and aviation, combined with slower-than-expected EV adoption, is driving this revision.

For oil and gas operators, one reality is clear: remote sites are not going offline anytime soon — in fact, their numbers will be growing in future. As a result, the importance of reliable data transfer from oil rigs to data centers continues to grow alongside the increasing volume of data. This capability opens such use cases as AI-driven predictive maintenance and real-time production optimization, unlocking greater efficiency, reduced downtime, and improved decision-making.

Remote operational environments expose the limits of legacy data infrastructure. When connectivity fails, data is lost or delayed. When it returns, systems are flooded with stale or duplicate information. Without prioritisation and backlog control, network recovery becomes the point of greatest operational risk.



In edge-to-cloud systems, the moment connectivity returns is often the moment failure begins.

What Is Not Working Today

Remote operational sites — drilling rigs, wellheads, field equipment — depend on satellite connectivity as their primary, and often only one, network connection. This creates a category of problems that most legacy data infrastructure was simply never designed to handle.

1. Data Loss and Stale Backlogs

When connectivity drops, most systems take one of two approaches: retry a fixed number of times and throw an error (losing data) or retry indefinitely (hanging or crashing the application). Neither is acceptable in an environment where a single missed sensor reading can affect safety or operational decisions.

When connectivity restores, the backlog of undelivered messages begins processing all at once. The result is a flood of stale, potentially duplicate data — sensor heartbeats sent 100 times during a 10-minute outage, for example — that overwhelms downstream systems and pollutes datasets relied upon by analytics platforms.

2. The Backpressure Effect

The most dangerous and overlooked failure mode in edge-to-cloud data systems is not the outage itself — it is the reconnection. When multiple remote edge applications simultaneously regain connectivity and attempt to transmit their accumulated backlogs, control systems are suddenly overwhelmed by a surge of traffic they are not designed to handle. The result is system instability, crashes, and data inconsistencies such as loss or duplication.

Alternatively, organizations are forced to overprovision infrastructure to absorb these spikes, leading to unnecessary operational costs.

In effect, the very moment the network recovers becomes the point of failure.

Legacy architectures have no answer to this problem. There is no prioritization, no rate control, and no intelligent backlog management.

3. Legacy Offload Platforms Cannot Scale

Several of our clients came to gravity9 after encountering performance and data transfer challenges with platforms built on Cassandra or proprietary technologies. These approaches consistently fail for the same reasons.

Operational data from edge sites is rarely relational. Forcing it into a fixed schema makes the data nearly impossible to query, maintain, extend or clean up.

One of our clients operated a Cassandra cluster with approximately 15 TB of active data, periodically spiking to 70–80 TB. A significant portion of this growth consisted of orphaned data that could not be effectively managed or cleaned up due to insufficient data isolation.

4. Proprietary Sync Solutions Are Being Deprecated

A significant portion of the industry relies on third-party mobile and device synchronization platforms for offline-first data collection at edge sites. These platforms abstract away the complexity of bi-directional sync, conflict resolution, and offline queuing. However, as vendors consolidate and deprecate older products, companies are left exposed — with dozens of applications dependent on a sync mechanism that is being switched off, and no clear migration path.

One of our clients had over 100 applications — mobile and desktop, spanning ~30,000 field devices — built on a synchronization platform scheduled for end-of-life. Rebuilding or replacing that infrastructure without disrupting ongoing field operations is a challenge that requires both architectural clarity and execution speed.

5. No Prioritization — All Data Is Treated Equally

Not all data from a remote site carries equal urgency. A safety system alarm is not the same as a routine equipment heartbeat. A drilling parameter update is not the same as a low-frequency administrative log. Yet most data transport systems treat every message identically — first in, first out — without any concept of business criticality.

This means that during constrained bandwidth conditions, critical operational data waits behind low-value telemetry. When the connection is limited, that ordering matters enormously.

Reliable edge-to-cloud data transfer isn't a solved problem it's the foundation every digital initiative depends on.

Our Solution

At gravity9, we believe that reliable edge-to-cloud data transfer is not a solved problem in the oil and gas industry — it is the foundational challenge that determines whether every other digital initiative succeeds or fails.

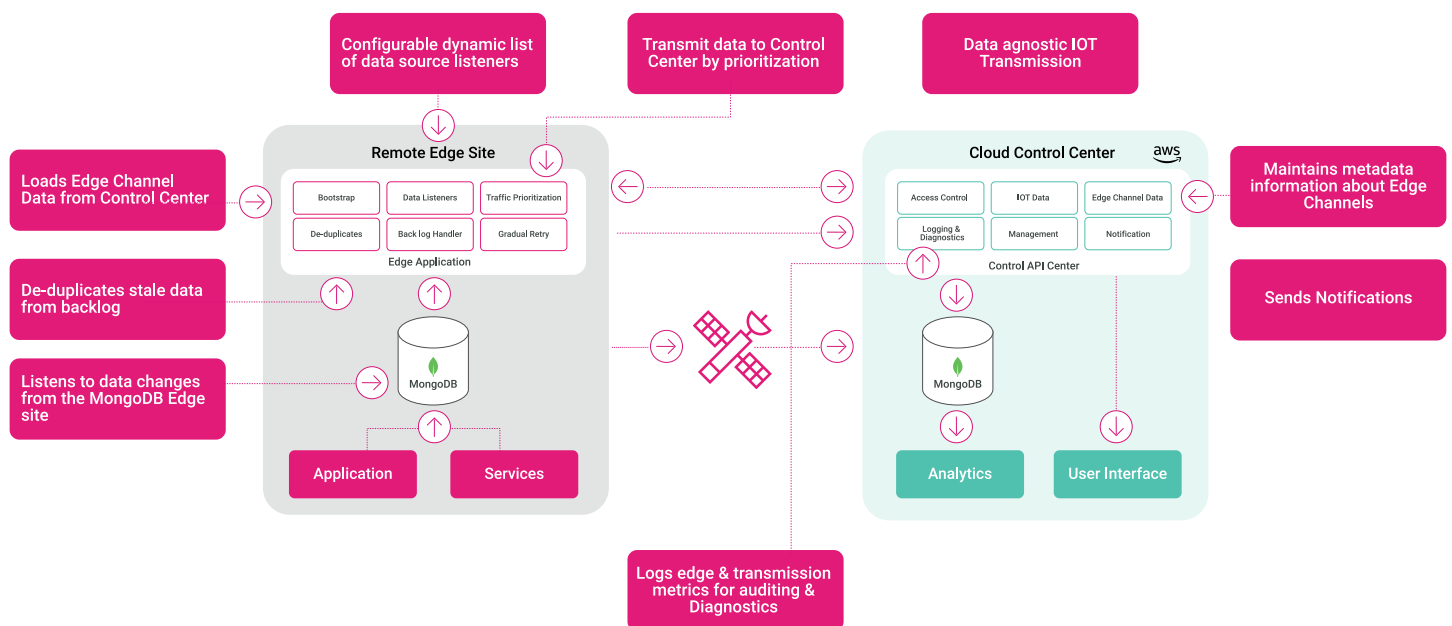
AI models are only as accurate as the data they train on. OSDU platforms are only as useful as the datasets they receive. Safety systems are only as reliable as their data feeds.

We also believe that solving this problem requires more than a point of integration. It requires a reusable, configurable framework that can be deployed across different remote sites, different applications, and different data volumes — without rebuilding from scratch each time.

The Medallion Edge Transmission Framework

The Medallion Edge Transmission Framework represents gravity9's vision for a reliable, prioritized data transfer solution between remote edge environments and centralized cloud control planes. Built on MongoDB and AWS, it is purpose-designed to address the unique connectivity challenges, data volumes, and operational constraints of the oil and gas industry, leveraging gravity9's practical experience in solving these complex, real-world problems.

The framework consists of two components that work together:



Edge Application (deployed on the remote site)

The Edge Application is installed on remote sites and operates as an offline-first data collector and transmitter.

On startup, it bootstraps itself from the Control Center, retrieving its full configuration: which data sources to listen to, how to prioritize them, what batch sizes to use, and where to transmit. This makes each Edge Application dynamically configurable without code changes.

When the connection to the Control Center is available, data flows in near real-time. When connectivity drops, the Edge Application aggregates data locally — using MongoDB applying deduplication logic to prevent stale data buildup. No data is lost, and no data is sent twice.

When connectivity restores, the Edge Application does not flood the control plane. Instead, it transmits data in controlled, prioritized batches according to its configuration and depending on available resources on the control center side.

Control Center (deployed in the cloud)

The Control Center is the cloud-side counterpart. It manages the configuration of every connected Edge Application, receives and persists incoming data, handles conflict resolution, and can push updates or corrections back to edge devices.

It is built as a microservice architecture on AWS, enabling horizontal scaling to handle increasing numbers of connected edge sites without architectural changes.

The Medallion prioritisation model ensures that even under constrained bandwidth, critical operational data is transmitted first, keeping systems responsive and decisions reliable.



Data Prioritization: The Medallion Model

Central to the framework is a configurable data prioritization model. Each data source at the edge is assigned a priority tier — example Gold, Silver, and Bronze as defaults, but priority definitions are fully configurable.

When transmitting a backlog under bandwidth constraints, the Edge Application composes each outgoing batch according to the configured priority ratio — for example, 60% Gold, 30% Silver, 10% Bronze. Critical operational data always moves first. Low-priority telemetry fills remaining capacity.

This means that even during degraded connectivity, the data that matters most reaches the control plane first.

Bi-Directional Data Transfer

The framework supports bi-directional data transit. While the dominant flow is from remote sites to the cloud — high-frequency telemetry, sensor readings, operational logs — the Control Center can push data back to the edge devices at a lower frequency.

This enables use cases such as:

- Sending updated calculation results or corrections back to field equipment
- Pushing new configuration parameters to edge devices remotely
- Delivering commands or alerts from Control Center to Edge Applications

Intelligent Edge: AI-Driven Optimization

The Medallion Edge Transmission Framework includes an AI layer that transforms the Control Center from a passive data receiver into an intelligent operations hub.

- Real-time utilization analysis: The AI continuously monitors each Edge Application's ingestion rate, connectivity patterns, and transmission metrics.
- Auto-tuning: Based on observed behaviour, the AI recommends and applies optimal transmission parameters — such as maximum batch size — without requiring manual intervention.
- Conversational interface: Operations teams can query the system in natural language: ask which edge applications are active, request current metrics, review ingestion logs, or instruct the system to update configuration parameters.

- Dynamic priority adjustment: The AI can adjust data prioritization in real time based on changing operational conditions — elevating the priority of sensors showing anomalous readings, for example.

Proof Points

A Global Oilfield Services Company — Device Sync at Scale

A leading oilfield technology company operates over 100 mobile and desktop applications across approximately 30,000 field devices worldwide, relying on a proprietary sync platform that is being deprecated. With the deadline approaching, they needed a replacement architecture they could own, extend, and reuse across their entire applications.

gravity9 designed and is delivering a custom bi-directional sync data pipeline built on Azure SignalR, Azure Functions, and MongoDB Atlas. The solution supports configurable conflict resolution, structured observability logs, and is designed for portability across both mobile and desktop form.

A Global Drilling Contractor — Pilot Data Transfer System

A major international drilling contractor operates remote well sites with unreliable data connectivity and a need to reliably transport operational telemetry to a centralized cloud platform. Their requirements included support for payloads of up to 1,000 data attributes per event, transmission frequencies of up to 1 update per second, bi-directional sync for periodic cloud-to-edge data, and a conflict resolution strategy for duplicate handling.

gravity9 is delivering this solution runs on MongoDB at the edge, MongoDB Atlas at the cloud control plane, and is built on the Microsoft Azure / .NET stack to match the client's existing technology investments.

A Global Energy Leader — Medallion data prioritization model

A global energy leader operating hundreds of remote drilling sites. Replacing the deprecated system built on Cassandra cluster by introducing a Medallion data prioritization model (Gold / Silver / Bronze), the client gained fine-grained control over what data transfer first — guaranteeing real-time visibility into mission-critical operations without saturating constrained bandwidth.

Gravity9 validated the full edge-to-cloud data pipeline using MongoDB Atlas and Confluent Kafka — setting the foundation for a production-grade, scalable solution.



Zero data loss.
Controlled recovery.
Critical data
delivered first.

Expected Outcomes

Operational

- Zero data loss: during connectivity outages — data always persisted locally until confirmed delivered
- Backpressure Effect: controlled, prioritized reconnect prevents overloading of control side
- Critical data always moves first: priority tiers ensure safety and operational data is never delayed behind low-value telemetry
- Operations centers receive consistent, timely data: enabling accurate decisions and reliable downstream platforms

Business

- Accelerated delivery: reusing the proven Medallion solution compresses development timelines.
- Reduced infrastructure cost: replacing over-engineered legacy data transfer platforms with purpose-built, scalable architecture
- Future-proof: framework ownership means no dependency on third-party sync vendors or proprietary platforms being deprecated

Why gravity9

We are MongoDB's Global Modernization Partner of the Year 2025, with 150+ engineers and a track record of delivering working solutions — not just architectural diagrams — for enterprises operating in demanding, data-intensive environments. Our approach in oil and gas is grounded in what we have observed across multiple engagements.

Visit our [Insights page](#) for more articles about emerging technology trends, the IT industry, interviews, and more!

Visit our [Insights page](#) for more articles about emerging technology trends, the IT industry, interviews, and more!

gravity9