crystal ∑quation

Crystal Equation Managed Services: DataOps

Over the past decade, Crystal Equation engineered and operated one of the most advanced DataOps ecosystems in the industry. The ecosystem evolved from manual oversight into a globally scaled, automation-driven operation spanning EMEA, APAC, and the US. The DataOps service covered more than 95% of the client's analytical data warehouse and delivered consistent reliability for mission-critical pipelines powering analytics, ML models, and product decisions.

Automation resolved over 75% of operational issues and 42% end-to-end without human involvement. In addition, automation prevented multi-million-dollar revenue impact by enabling partition level dependency tracking across the warehouse. This resulted in a resilient, proactive, and largely autonomous DataOps capability that minimizes the need for manual intervention while enabling engineering teams to focus on innovation.

Challenges the Client Initially Faced

- The client's fast-growing data warehouse lacked the foundational processes and tooling needed for reliable operations. Critical pipelines frequently arrived late or failed without visibility. In addition, there was no unified system to track delays, dependencies, or pipeline health. Data engineers spent a significant portion of their day manually troubleshooting issues because no consistent triage process, RCA model, or operational framework existed.
- Each team worked in isolation, using different tools and approaches, which led to inconsistent quality, repeated failures, and no horizontal view across the warehouse. With no standardized DataOps practice in place, operational gaps continued to grow, increasing the risk of incorrect insights and delayed business decisions.
- The client needed a partner who could introduce global standards, build the missing tooling, and create an automated, reliable DataOps layer capable of supporting the warehouse at scale. They needed a team willing to dive deep into the ecosystem, break it down to its core components, and rebuild it using best practices, modern automation, and purpose-built tooling.

Approach to the Solution

Deep discovery through manual operations	We began by hiring experienced software engineers and immersing them directly in day-to-day operational work across the data warehouse. For months, they manually triaged failures, mapped dependencies, and investigated pipeline issues to deeply understand the pain points. This "learn by doing" approach created unmatched insight into failure patterns and allowed us to design tools and automation based on real operational needs — something never attempted at this scale.
Full audit & ecosystem mapping	• Comprehensive review of pipelines, SLAs, owners, and failure hotspots • Identification of systemic risks, bottlenecks, and missing processes • Complete dependency and data-flow mapping across regions
Building the global DataOps framework	We established the foundational processes the client lacked: • Standardized triage playbooks, escalation paths, and root-cause analysis models • Unified global communication templates • Daily cross-regional syncs to create alignment between US, EMEA, and APAC • Horizontal, warehouse-wide visibility to move away from siloed operations

Rapid tooling development & automation	Because our engineers were building tools they themselves used, iteration and adoption were extremely fast. We delivered: • Automated monitoring, dependency checks, and anomaly detection • Alert deduplication and early failure detection • Automation of repetitive tasks and shift-left workflows • Internal platforms now foundational to the client's data infrastructure
Scaling down human involvement through automation	As automation matured, SWE expertise became less critical for day-to-day operations. This allowed us to transition from high-cost engineering staffing to a lean support model: • **Reduced from 42 specialists across 6 regions** • **Down to 14 specialists in 2 regions** • **While supporting 10× more workload** The system is now on track to fully remove human-in-the-loop for most operational tasks by end of 2026.
Close partnership with engineering teams	Throughout the journey, we collaborated with client infrastructure teams to integrate our tools into core systems, improve resiliency, and feed insights back into long-term engineering roadmaps.

Solution

We delivered a fully managed, globally scaled DataOps solution that transformed the client's fragmented, high-effort operational landscape into an automated, efficient, and predictable system. Over nearly a decade, the service grew from a small manual function supporting one customer into a robust ecosystem covering 95.5% of the analytical warehouse, supporting thousands of internal users, and handling 10× the original workload with a team reduced from 42 specialists across six regions to 14 across two.

Our solution combines three integrated pillars — operational reliability, analytical visibility, and deep automation — enhanced by a suite of purpose-built products.

Operational Layer

Reliable, global 24/7 coverage

- Continuous monitoring of 1M+ critical pipelines across EMEA, APAC, and US
- Automated dependency checks and delay detection before downstream impact
- SLA-driven triage ensuring consistent and rapid response
- Follow-the-sun operations with unified practices and cross-regional consistency
- A lean model: 42 → 14 specialists while supporting 10× more volume

This operational foundation established reliability, consistency, and predictable delivery.

Engineering Layer

Full visibility into pipeline health

- **OpCenter: Centralized operations console.**
  - A one-stop interface for managing pipeline delays and failures.
  - Consolidates troubleshooting information
  - Streamlines escalation and remediation
  - Reduces on-call time and effort
- **Kayak: Dependency graphing interface**
  - A visual system that maps dependencies between datasets.
  - Automatically identifies upstream blockers
  - Eliminates manual and time-consuming investigation
  - Accelerates root-cause discovery
- **Dependency service & analyzer service**
  - Core infrastructure services enabling real-time dependency traversal and pipeline health insights across the warehouse.

These tools provided the horizontal, warehouse-wide clarity the client previously lacked.

Automation Layer

A system that runs itself

Our automation framework evolved into the backbone of DataOps, enabling rapid remediation and proactive detection:

- **Auto-remediation framework**
  - Automates resolution for failures with known patterns.
  - Handles recurring "low-hanging-fruit" issues
  - Allows specialists to focus on complex problems
  - Reduces manual intervention dramatically
- **Early Warning System (EWS) — Our flagship automation**

- A predictive engine that determines if a critical pipeline will miss its SLA.
- Sends early delay notifications
- Enables proactive fixes before impact
- Prevents major severity incidents

With automation resolving 75%+ of issues and 42% fully handled without human involvement, the system is now on track for full human-out-of-the-loop DataOps by 2026.

To summarize, this solution didn't just solve the client's challenges. It rebuilt and redefined DataOps

into an automated, scalable, and future-ready ecosystem that continues to deliver warehouse-wide reliability with minimal human intervention.