WHITEPAPER

Agentic AI in Action EXPLORE

The Evolution of the Data Engineer

Rethinking pipelines, data operations, and data quality for the Agentic AI era



Qlik

qlik.com

The pipeline is only part of the story.

As data engineers, you know AI-driven systems are demanding more. A lot more, like:

- More context to guide AI reasoning
- More visibility into how data moves and changes
- More confidence in the quality and trustworthiness of data
- More resilience as systems, workloads, and dependencies evolve

Much of the conversation around AI in data engineering has focused on speeding up development through coding agents, declarative workflows, and AI-assisted pipeline creation.

Those advances matter.

The bigger shift is operational.

Modern AI systems require more than faster development. They depend on context, trust, and operational intelligence that can move with data as conditions evolve.

That means rethinking three core aspects of modern data engineering:

1. How pipelines are built
2. How operational context supports decision making
3. How trust is maintained at runtime



Pipeline development becomes intent driven.

AI-driven systems are expanding how pipelines get built and maintained. While automation has long been part of modern data engineering, the introduction of coding agents, natural language interfaces, and declarative workflows is giving engineers more options than ever for how they develop and orchestrate data workflows.

Whether the priority is speed, control, flexibility, or productivity, engineers now have more ways to get the job done.

The market is shifting from pipeline building — which relies on traditional, process-driven workflows — to system orchestration, which requires modern, intent-driven workflows.

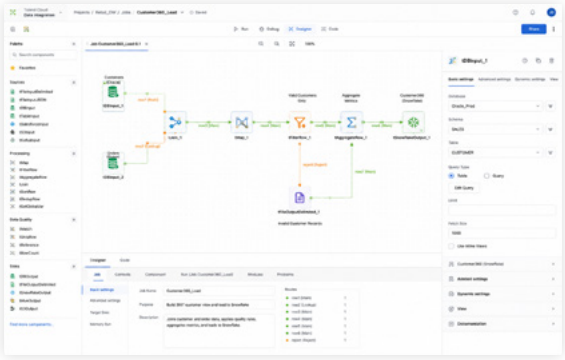
Here's how this operating model is evolving:

PROCESS-DRIVEN WORKFLOWS

To build and maintain pipelines, data engineers have always needed to know not only what to build, but how to build it.

- Step-by-step configuration
- Manual orchestration
- Predefined logic
- Lower-level implementation details

Automation has helped streamline parts of the above process, but engineers still define workflow logic, orchestrate pipeline behavior, and manage implementation details. Process-driven workflows provide structure and control through explicit configuration and predefined logic.



Engineers define both the what and the how.

VS

INTENT-DRIVEN WORKFLOWS

Natural language prompts, AI-driven pipeline generation, and declarative orchestration are reducing the amount of manual effort required to build and maintain modern data workflows.

Instead of manually configuring every pipeline step, engineers define the desired outcome while AI-assisted systems help accelerate implementation behind the scenes.

This shifts pipeline development toward:

- Outcome-oriented workflows
- Intent specification
- AI-assisted implementation
- Reduced manual configuration

Engineers can build pipelines using coding agents like Claude Code, GitHub Copilot, and natural language AI assistants.

Build a pipeline that combines customer transations, support tickets, and product telemetry to identify churn risk in near-real time.

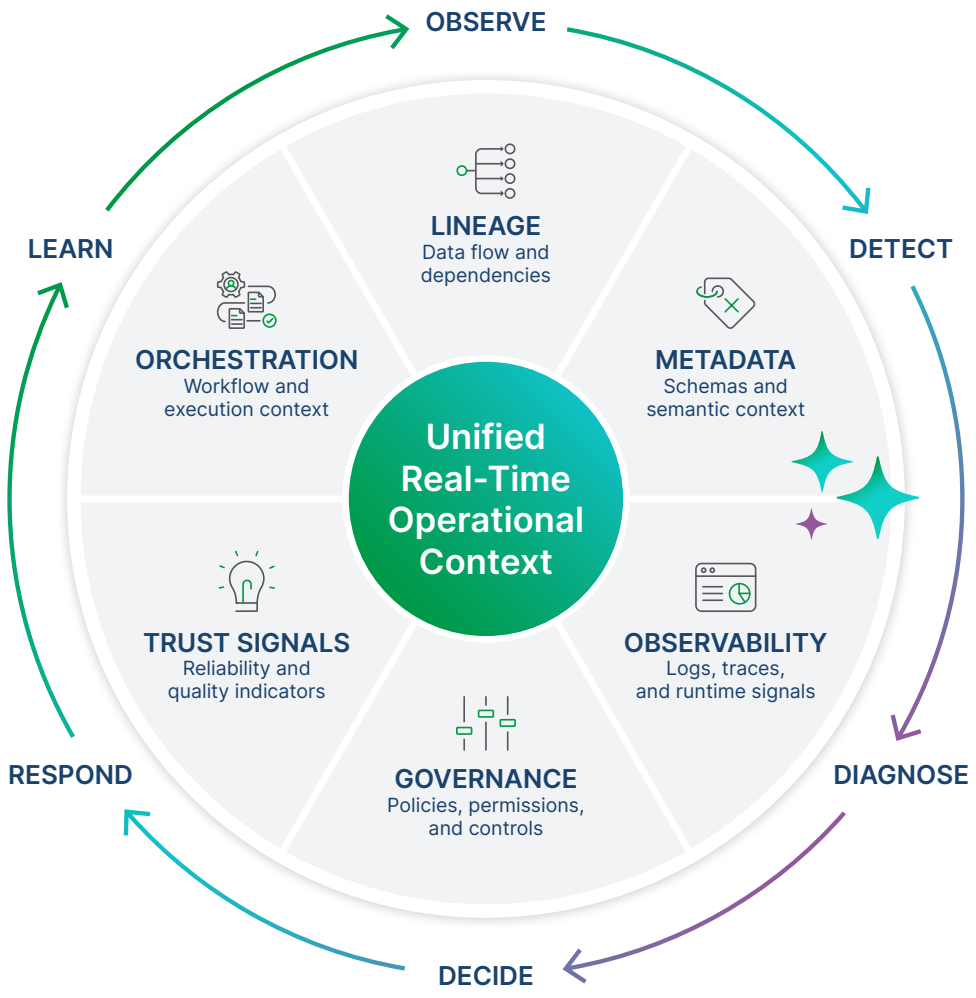
Engineers define the what while AI-assisted tools help accelerate the how.

Operations closes the context gap.

Building pipelines is only half the challenge. Keeping them resilient as systems, workloads, and dependencies evolve is where the real complexity begins.

Modern operations generate no shortage of signals. The hard part is turning those signals into usable context.

Continuous Intelligence Loop



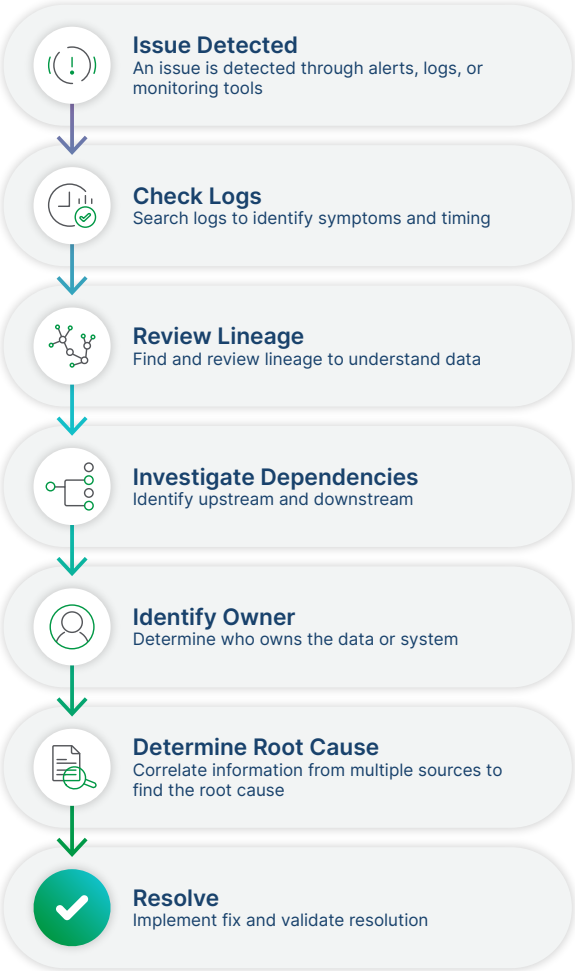
Unified real-time operational context helps teams and AI systems:
Observe, understand, and respond with greater speed and confidence

Less time gathering context.
More time solving the problem.

When issues arise, the challenge isn't finding data — it's finding the operational metadata, lineage, and background context needed to understand what happened and what to do next. Connecting that context helps teams spend less time investigating data issues, resolve problems faster, and operate more efficiently at scale.

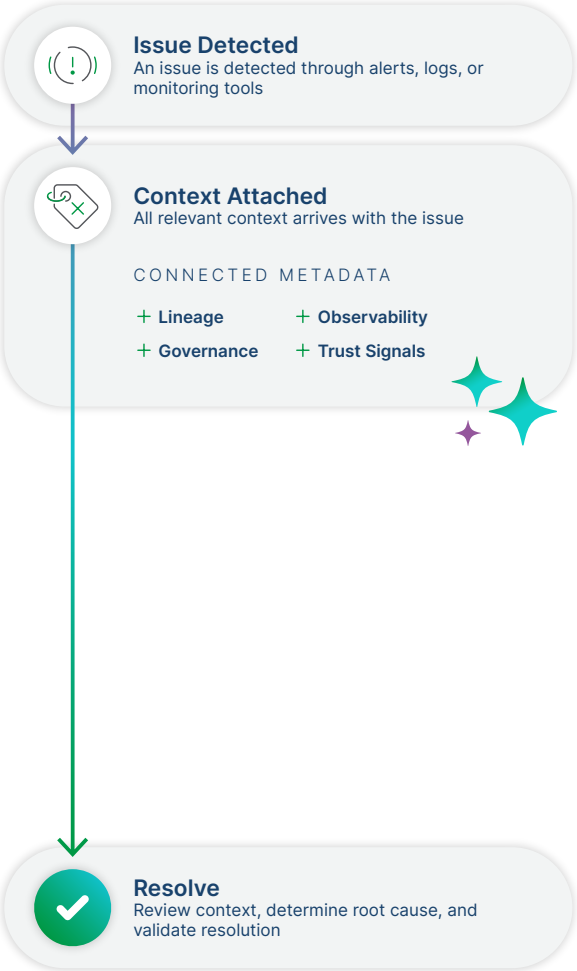
TRADITIONAL DATA OPERATIONS

Issues are often surfaced through disconnected alerts, logs, and monitoring tools. Engineers must manually gather lineage, dependencies, ownership information, and other operational metadata across systems before determining the root cause and resolving the problem.



CONTEXT-AWARE DATA OPERATIONS

Metadata, including lineage, observability, governance, and trust signals, remains connected to the workflow. When issues surface, the context needed to understand and resolve the problem is available immediately for human judgment to diagnose and implement the validated fix.



Recommendations can be automated. Accountability can't.

Context-aware data operations help reduce investigation time and accelerate remediation. But faster decisions don't eliminate the need for human judgment. Engineers play an even more critical role in reviewing recommendations, approving actions, and overseeing complex operational workflows.

As more tasks become automated, engineers become increasingly responsible for providing oversight, applying judgment, and maintaining control.

OVERSIGHT

Ensure trust and accountability through governance and controls

JUDGMENT

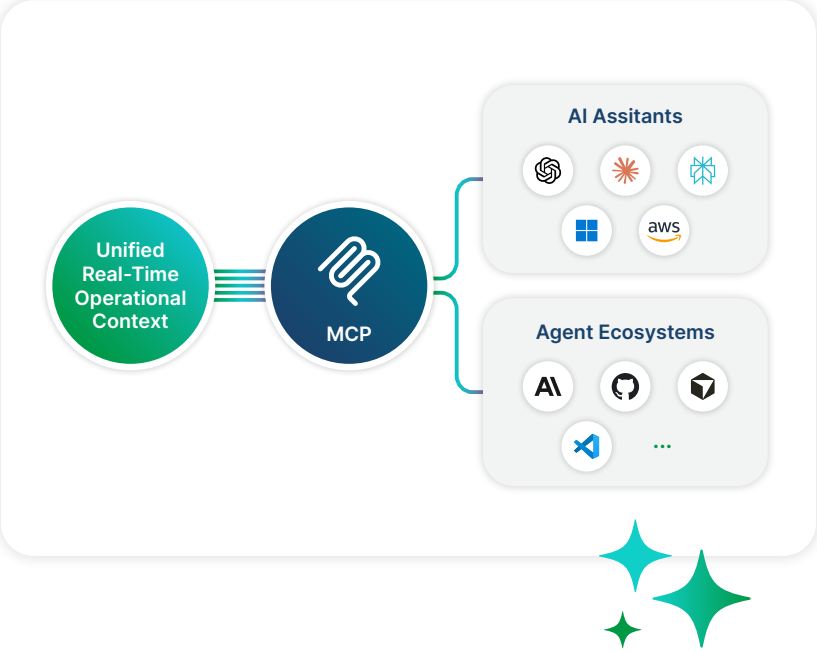
Apply human expertise to evaluate context, risk, and impact

CONTROL

Approve, adjust, or override recommendations when needed

Extending context across systems.

With Model Context Protocol (MCP), data platforms can extend trusted context and robust governance directly to AI assistants and connected agent ecosystems. The seamless interoperability between disparate data sources and AI models ensures intelligent systems respond with greater accuracy, consistency, and reliability.



Context changes everything.

AI systems can only act on the context available to them. That's why operational context needs to move with the workflow — not sit in disconnected dashboards, documentation, or metadata repositories.

When lineage, runtime conditions, governance policies, data quality, and dependencies are continuously connected, teams and AI systems can respond with greater speed, accuracy, and confidence.

This is the shift facing today's data engineers, from building pipelines to enabling systems that can operate with greater context, trust, and intelligence.

Trust becomes operational.

Trust can't be built on static governance policies and documentation. Modern data operations need real-time signals that help determine whether data is trustworthy, current, and safe to act on.

- **Data reliability**
- **Lineage visibility**
- **Governance controls**
- **Quality conditions**

Without this foundation, AI systems struggle to reason consistently, increasing the risk of hallucinations, remediation errors, and downstream failures. And as agentic workflows scale, those failures become harder to isolate across connected systems and processes.



Signals like Qlik Trust Score® help teams continuously evaluate whether data is reliable, governed, current, and ready for AI workflows, giving agents greater confidence as they reason, recommend, and act.

What makes data ready for agentic workflows?

Not every dataset is ready for AI-driven workflows. Governed data products embed semantics, lineage, governance, and quality signals so agents can reason with full context, not just data. This helps ensure AI-driven workflows operate on trusted, governed, and current data with greater consistency and reliability.

AI agents succeed or fail on data quality. As organizations increasingly rely on AI-driven workflows, trusted data products help ensure agents reason against governed, current, and context-rich data — not incomplete or outdated information.

Four qualities of a trusted data product

1. **Traceable lineage**

   Dependencies, provenance, and system impact can be traced across workflows and environments.

2. **Shared definitions**

   Consistent business logic and meaning are maintained across workflows and environments.

3. **Inherited governance and access controls**

   Policies, permissions, and validation rules are consistently applied so agents operate with built-in accountability and clear guardrails.

4. **Data freshness**

   AI systems are operating against current runtime conditions.

The role of the data engineer is changing in exciting ways.



New AI systems require more than faster pipeline development or smarter automation. They depend on resilient infrastructure, trusted data, operational context, and workflows that can adapt as conditions evolve.

Velocity from AI. Vision from engineers.

By reducing operational friction and accelerating development, engineering teams can spend more time designing resilient systems, governing trusted data, and building the foundations that AI workflows depend on. With the right tools, like agents built for the builders, AI may accelerate execution, but engineers remain responsible for the architecture, governance, and decisions behind it.

Discover agentic data engineering and build with intent.

Explore Now

Qlik

Agentic AI in Action EXPLORE

About Qlik

Qlik helps teams get more out of AI with data they can rely on and control. It brings trusted data products, a powerful analytics engine and AI agents together to reduce risk, keep operating costs in check, and scale AI responsibly as needs evolve. Used by 75% of the Fortune 500, Qlik supports customers worldwide. Qlik works with the systems and partners customers already use, so teams can stay flexible without lock-in.

qlik.com

© 2026 QlikTech International AB. All rights reserved. All company and/or product names may be trade names, trademarks, and/or registered trademarks of the respective owners with which they are associated. For the full list of Qlik trademarks, please visit: https://www.qlik.com/us/legal/trademarks

26-AI-2610-02-WHM