Artificial Intelligence

From weeks to minutes: How Formula 1® uses agentic AI on AWS to accelerate data operations

by Subhro Bose, Arunraja Kumar, Jerome Descreux, and Matt Kemp | on 03 AUG 2026 | in Amazon Bedrock AgentCore, Amazon SageMaker Unified Studio, Artificial Intelligence, Customer Solutions, Intermediate (200), Technical How-to | Permalink | Share

Formula 1® (F1) engages an audience of over 800 million fans globally across digital platforms, F1 TV, social media, ticketing, and merchandise year-round. Races happen every two weeks. Fan engagement windows are measured in minutes and commercial decisions need to move at the speed of the grid. Behind the scenes, F1's marketing technology (MarTech) platform, Customer 360, captures interactions across all of these touchpoints to power personalization, segmentation, and commercial strategy.

However, the platform faced a significant operational challenge. According to Chris Roberts, Director of IT at Formula 1, "Our MarTech platform is the nervous system of F1's fan engagement. But every new data source required 6 to 8 weeks of manual engineering. We had an 18-month backlog just to integrate 12 new sources." The business was generating data faster than the engineering team could wire it up. As a result, Matt Kemp, F1 Head of Data Operations, set to improve efficiencies and data quality. "Manually ingesting data sources is time consuming, creates solution variances, and ultimately results in data integrity issues. I wanted a solution that was repeatable, robust and reliable. AWS worked backwards from our needs to implement an agentic solution that worked end to end, applying business logic at each step."

In early 2026, F1 and AWS worked together to build the Data Accelerator, a solution that uses agentic AI on Amazon Bedrock AgentCore to transform F1's MarTech data platform from a manually maintained system into a self-managed, observable, and unified data estate. In this post, we show how the Data Accelerator reduced data source onboarding from up to 8 weeks to approximately 40 minutes of code generation plus hours of deployment. It also identified and fixed data source anomalies in production, tracked data platform operations and agent lineage in a single window, and opened a gateway for analysts, engineers, and scientists to collaborate. "For the first time, we have end-to-end visibility across the entire MarTech platform with data lineage and root cause analysis, not just dashboards full of alerts," says Roberts.

The challenge

F1's Customer 360 platform ingests data from ticketing partners, streaming integrations, sponsor activation feeds, social media, and merchandise systems. Operating a data estate of this breadth and velocity surfaced three areas of friction the team set out to solve. First, onboarding each new data source was a heavily manual effort: engineers wrote schema mappings, built ingestion pipelines, configured data quality checks, defined General Data Protection Regulation (GDPR) classifications, and set governance policies by hand. This process took 6 to 8 weeks per source. Second, the platform had to keep pace with constantly evolving upstream feeds. Providers frequently changed column names, added fields, or restructured and rescheduled payloads without notice. Those changes often surfaced at the worst possible moment, such as mid race-weekend or during a mission-critical campaign launch. Third, visibility was fragmented. Logs were scattered across services with no unified data lineage. When a stakeholder questioned a metric, engineers spent hours manually tracing the issue across Amazon Simple Storage Service (Amazon S3) paths, Amazon Redshift control tables, Airflow logs, and DBT outputs.

Solution overview

The Data Accelerator addressed these challenges through five workstreams delivered simultaneously:

- Agentic data source onboarding using Amazon Bedrock AgentCore, hosting agents in its runtime containers.
- Automated schema evolution detection and remediation.
- Unified data access through Amazon SageMaker Unified Studio.
- End-to-end observability with root cause analysis tool (RCA) and context graph.
- Automated identification of a failure in observability dashboard and agentic operation if they could be fixed with code changes.

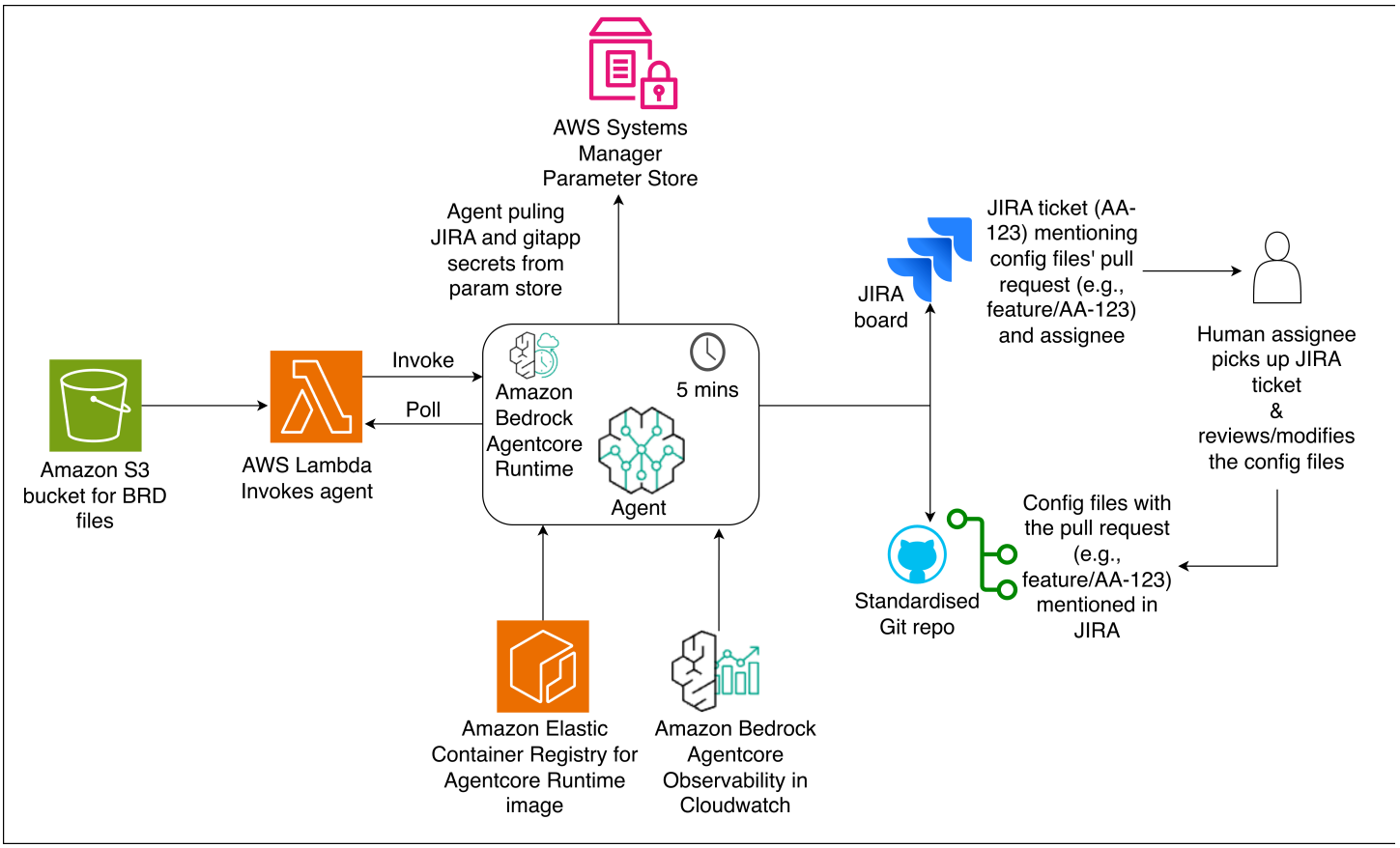
A sixth workstream optimized the customer identity resolution algorithms that unify fan touchpoints across channels. The following sections describe each workstream in detail.

Agentic data source onboarding

The centerpiece of the Data Accelerator is a set of platform agents that take a Business Requirements Document (BRD) with limited information about the data source and produce a fully production-ready onboarding pipeline. This includes infrastructure code, data transformations, governance policies, and GDPR classification without a human writing a single line of boilerplate. The agents work in two phases:

Phase 1: Configuration generation

When a new data source needs onboarding, a team member uploads a BRD to an Amazon S3 bucket. The upload triggers an AWS Lambda function, which invokes Amazon Bedrock AgentCore Runtime, a capability of Amazon Bedrock AgentCore. The agent reads the BRD and generates a set of configuration files. It then accesses GitHub through a GitHub App to push these files as a pull request to the standardized Git repository, and accesses Jira through its REST API to create a ticket referencing the PR. All agent conversations and actions are traced in Amazon CloudWatch through built-in AgentCore observability. The assigned engineer reviews, adjusts if necessary, and approves.



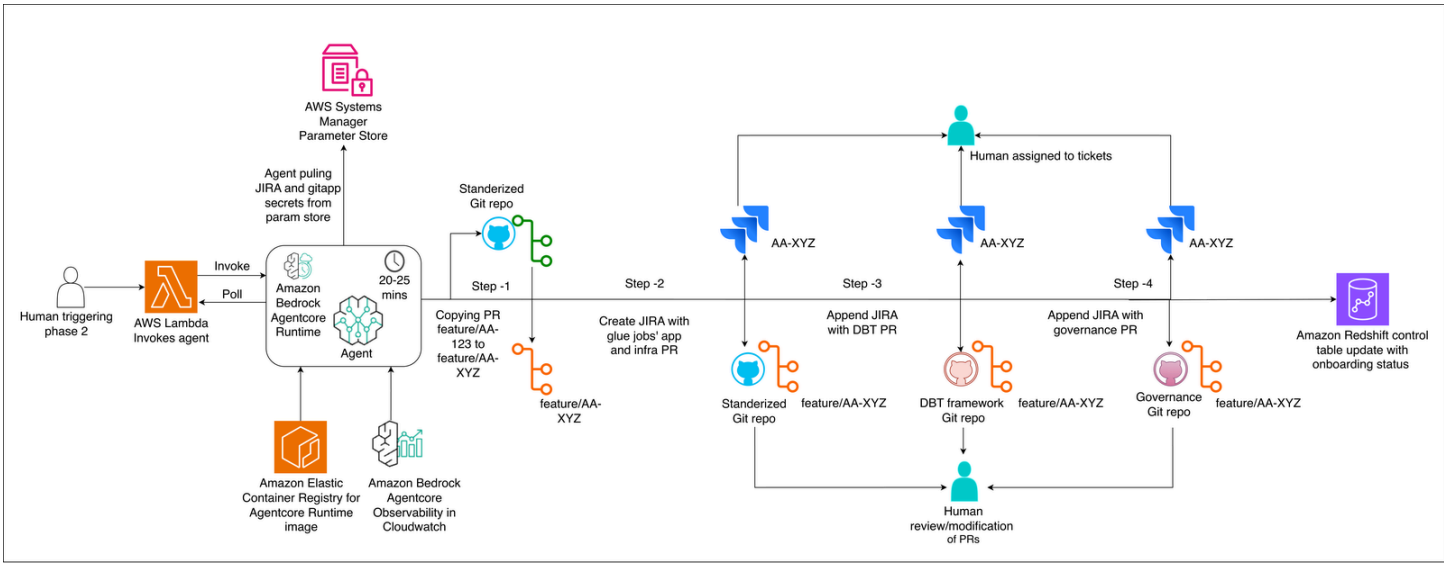
Phase 1 workflow: a BRD upload triggers the agent to generate config files and open a pull request

Phase 2: Full pipeline generation

Once the configuration files are approved, a human triggers the next stage. The agent takes the approved configuration and generates three separate Pull Requests:

1. AWS Glue application and infrastructure code.
2. DBT transformation framework.
3. Governance policies including GDPR tagging.

All three PRs link to a single Jira ticket for traceability. Engineers review each one across the Infrastructure, DBT, and Governance repositories and approve.



Phase 2 workflow: the agent generates infrastructure, transformation, and governance pull requests

Automated GDPR classification

What distinguishes this from a basic code generator is the integrated GDPR classification. The agent proactively analyzes every data column, determines whether it contains personal data, sensitive personal data, or pseudonymized data, and tags it with the appropriate GDPR category. These tags publish directly to the governance registry in SageMaker Unified Studio, giving the compliance team immediate visibility without manual review cycles.

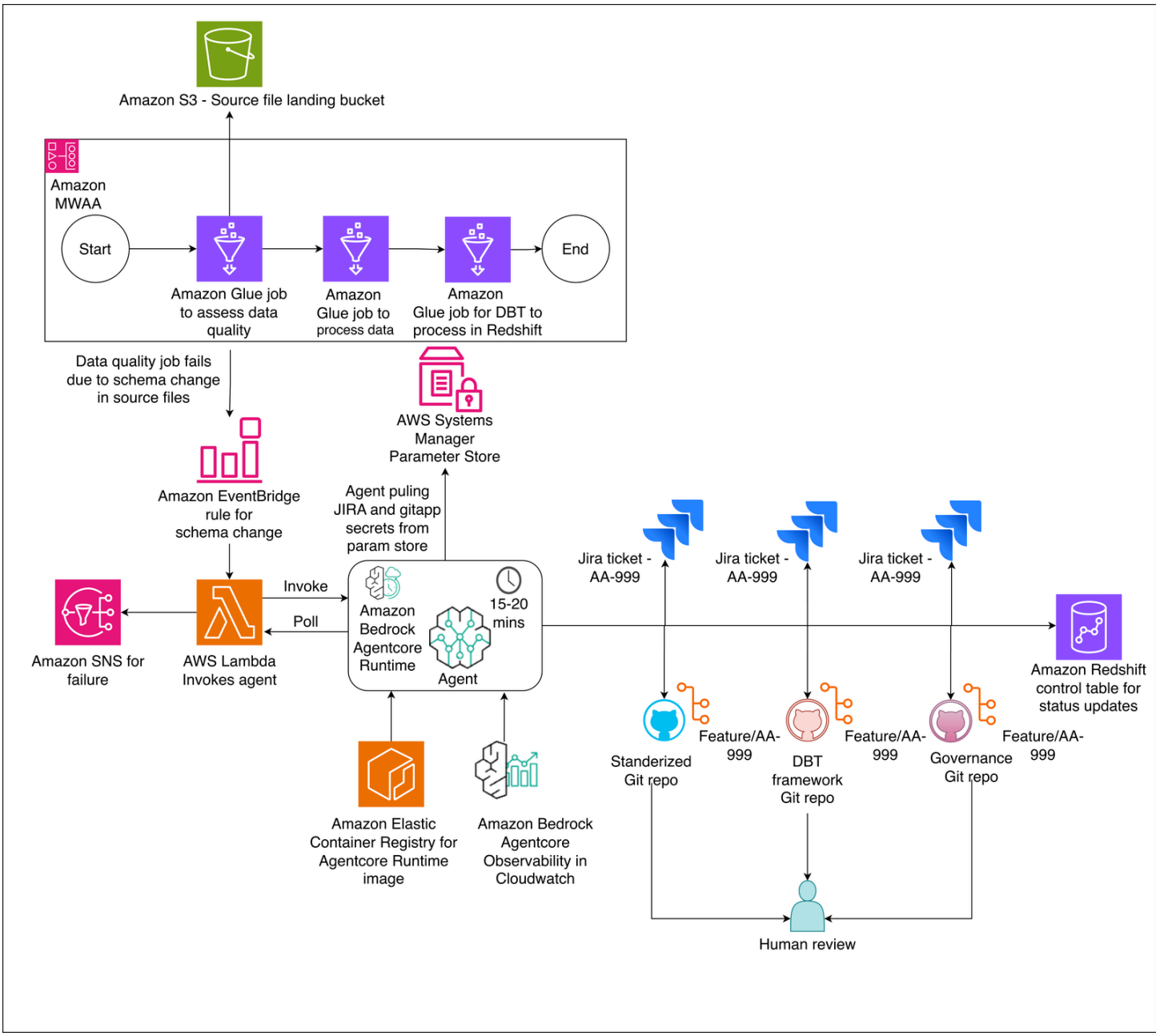
Modular skill architecture

The system is not a tightly coupled agent graph. A single agent operates with modular skill definitions, each encapsulating a distinct capability: schema mapping and data type inference, data quality validation, governance enforcement, and sensitive data classification. At runtime, the agent evaluates incoming requirements and activates the relevant skills, composing them through a multi-pass reasoning process. Pass-0 handles token management through scrubbing, Pass-1 summarizes tool outputs, and Pass-2 rolls up an overall assessment, refining accuracy and completeness progressively rather than relying on a one-shot response. New capabilities ship as new skill modules without changing the core agent loop, keeping the architecture maintainable and composable as the platform grows.

The result is onboarding time dropped from 6 to 8 weeks to approximately 40 minutes of code generation plus hours of deployment and review. AI agents now handle 95% of the work autonomously.

Automated schema evolution

Onboarding new data sources is one challenge, but keeping existing integrations healthy is another. Upstream providers frequently modify their data structures, from renaming a column to creating a new field. Previously, the F1 team discovered these changes when a pipeline failed, often during a live race weekend. The same agent architecture that handles onboarding now continuously monitors for upstream schema changes. When a provider modifies their data structure, the agent detects it through event-driven triggers using AWS Lambda and Amazon EventBridge. It assesses the downstream impact, identifying which pipelines are affected, and which consumers depend on the changed fields. It then generates the necessary code updates across all affected repositories and creates a Jira ticket with full context and linked PRs. Engineers receive a notification that explains what changed, describes the impact, and presents a proposed fix for review. End-to-end resolution now takes hours instead of days.



Schema evolution agentic workflow

Unified data access with Amazon SageMaker Unified Studio

Before the Data Accelerator, working with Customer 360 data required navigating multiple disconnected environments. Data engineers curated pipelines in one account. Data scientists who wanted to model fan behavior needed access to a separate account, and analysts operated in a third world entirely. Nobody shared tooling or context, and getting from a question to an answer took days of coordination before any analysis could begin.

The solution uses Amazon SageMaker Unified Studio as the foundation for a data mesh framework where a central governance account brokers data discovery and access across multiple producer teams. The key enabler: governance is codified as declarative configuration, not manual console operations. A single data source definition simultaneously publishes data to the catalog and provisions the access control needed for consumers to subscribe. This means agents can safely onboard new data products end-to-end, from storage to catalog to governed access, because the framework enforces security constraints by construction. No human needs to review

IAM policies or AWS Lake Formation grants. The platform guarantees correctness structurally. This is what makes the “one front door” possible.

Data engineers curate and govern datasets in one place, and data scientists find those same datasets in the same environment: governed, documented, and ready to model. A data scientist building a fan segmentation model or optimizing the customer identity algorithm doesn’t need to know where the data lives, who owns the pipeline, or which S3 prefix to use. They open Unified Studio, find the curated Customer 360 datasets, and start modeling. They get shared notebooks, consistent tooling, and governed access, because declarative governance made safe self-service possible without sacrificing control. The curation and the consumption finally live side by side.

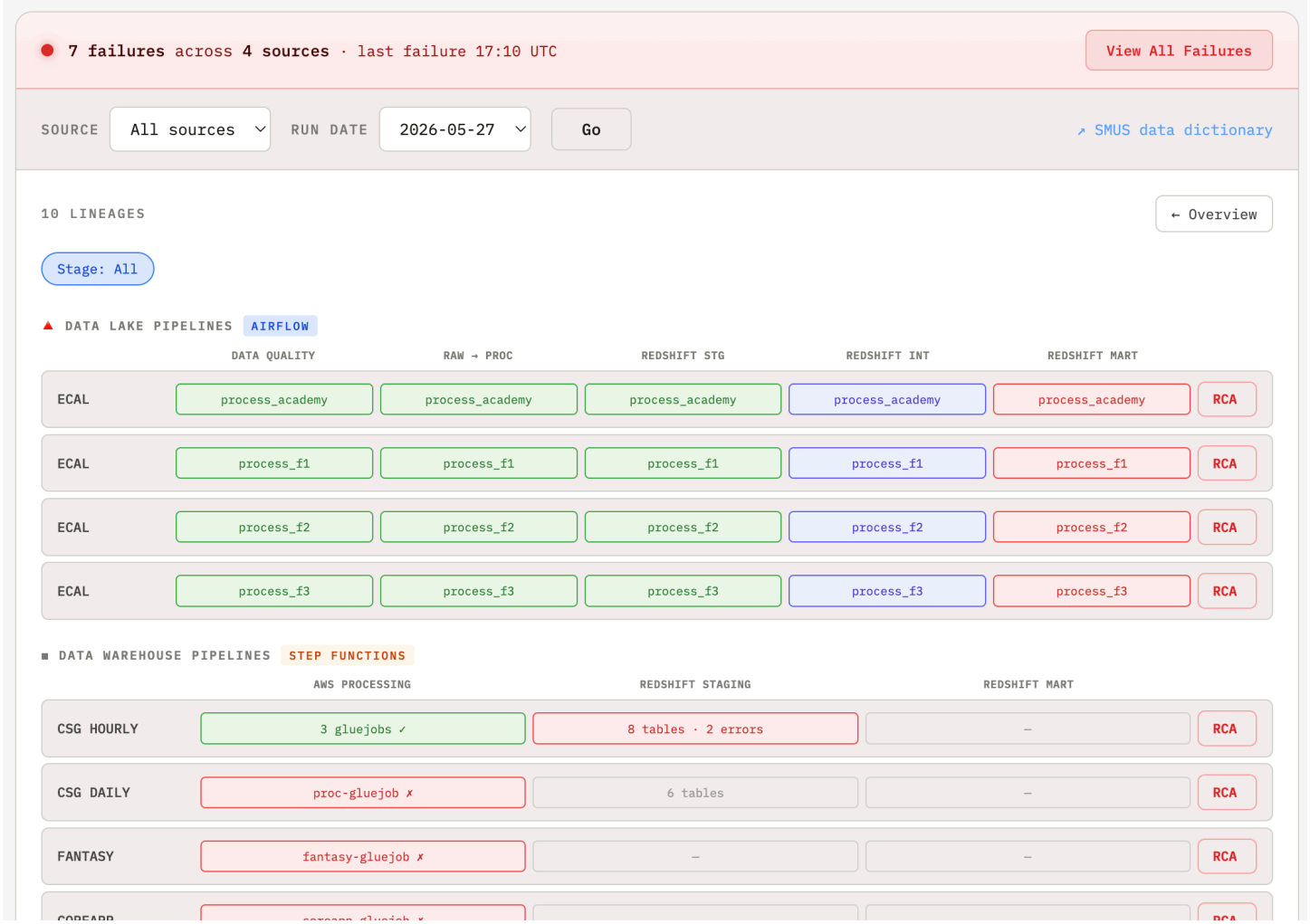
End-to-end observability with RCA and context graph

A data platform is only as trustworthy as the team’s ability to answer one question: is the data correct right now? Before the Data Accelerator, answering that question meant logging into Apache Airflow, checking Amazon S3 paths, querying Amazon Redshift control tables, and reading DBT logs. “Nobody had the full view. When a stakeholder asked, ‘why does this number look wrong?’ the answer was always, ‘give us a few hours.’ The observability dashboard changes that entirely,” adds Roberts.

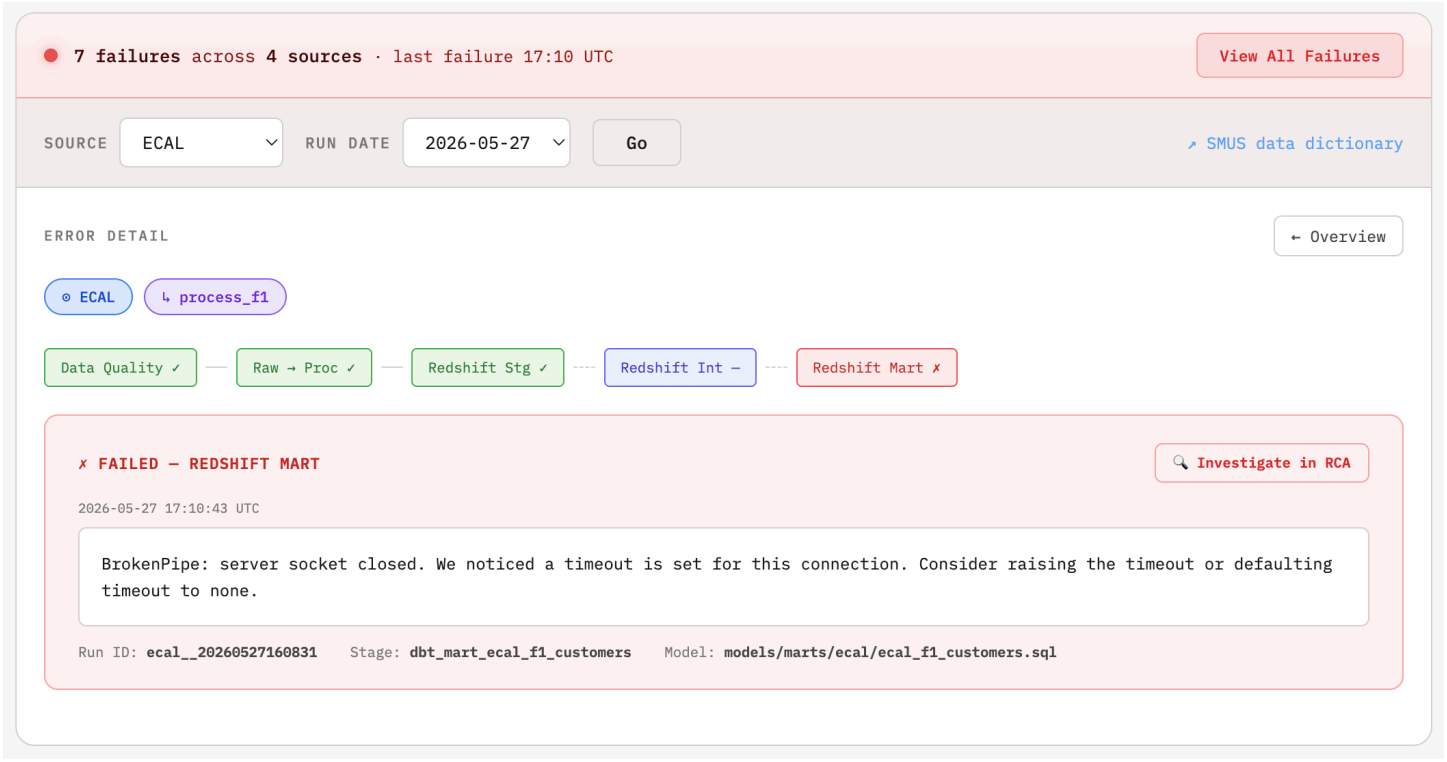
The observability layer presents full data lineage from S3 Raw ingestion through Processed layers into Amazon Redshift DBT stages as a single interactive graph, color-coded for health. Users click on any node to drill down to individual sources and tables, each showing pass/fail status, last run time, and duration. If a pipeline fails, the lineage visualization shows exactly where the break occurred, and which downstream data is affected.

Root cause analysis (RCA) is an agentic tool within F1’s platform that reads system logs and identifies failure points across the data estate. On its own, RCA can tell you what failed. We augment the RCA tool by passing through business context and system topology, codified as JSON. A missing file in S3 might be the error, but with the context graph, RCA tells you that the upstream provider rescheduled their delivery window, which is why the file wasn’t there when the pipeline ran. That’s the difference between knowing what failed and understanding why.

For the first time, F1 has full lineage, causal root cause analysis, and business context definitions in one place, and the dashboard auto-refreshes every 15 minutes.



Data lineage visualization showing pipeline health across sources and stages



Observability dashboard with failure details

Customer identity resolution

The final workstream optimized the algorithms that resolve customer identity across F1's fan touchpoints. A single fan might interact through the app, buy tickets on the website, watch on F1 TV, and engage on social media. Unifying those interactions into a single identity without false merges or missed matches is what makes effective personalization possible within the Fan Personalization Platform (FPP).

F1 already had a working identity resolution process, but it was slow and struggled to scale with the growing volume of fan interactions across channels. Rather than re-architecting the pipeline or replacing components, the team focused on optimizing the existing resolution algorithm's computational performance. By profiling execution bottlenecks and tuning the matching logic, the engagement reduced processing time by 50%, while keeping the entire resolution pipeline and its downstream integrations fully intact. No processes were changed, no accuracy trade-offs were made: the same algorithm now runs in half the time at F1's production scale.

With faster resolution, F1 can onboard any new data source and gather new customer data in half the existing time. Faster resolution means fresher unified profiles, which in turn means more timely and relevant personalization across every marketing channel.

"The whole point is to deliver the right message to the right fan at the right time, whether that's through email, the F1 app, ticketing, or social. Now that we can onboard sources in hours and resolve identities faster, we can actually deliver the personalized experiences our fans expect across every marketing channel," says Kemp.

Security and governance by design

The Data Accelerator operates on the principle that AI proposes and humans review. The agents run on Amazon Bedrock AgentCore with long-term memory, retaining context across invocations. Development used Kiro for structured spec-driven development and Amazon Bedrock (Claude Sonnet 4.6) as the foundation model. The event-driven backbone uses AWS Lambda for compute, Amazon EventBridge for routing, Amazon Managed Workflows for Apache Airflow (MWAA) for workflow orchestration, and Amazon S3 as the raw data layer. All AI model access is governed through F1's AI Gateway for unified access control, cost management, and audit logging. But the architecture is only half the story. The security posture is what makes this production-ready.

The security posture includes:

- Least privilege: fine-grained permissions, short-lived tokens with one-hour expiry, access limited to specific repositories and resources.
- Full audit trail: every action is logged and attributed for compliance.
- Human review: every generated Pull Request goes through engineer approval.
- Automated testing: agents generate comprehensive tests for their own changes.
- Rollback capabilities: issues surfaced post-merge can be reverted immediately.
- Network isolation: the entire system runs within private subnets in Amazon Virtual Private Cloud (Amazon VPC) with no direct internet access.
- Encrypted credentials: all secrets stored at rest in AWS Systems Manager Parameter Store.

"What gave us confidence to put agentic AI in our production data pipelines was what we call 'Human at the helm.' The agents do the heavy lifting, but humans make the decisions. Every change goes through the same review process our engineers already use, so adoption was immediate." says Roberts.

The impact

The Data Accelerator delivered measurable impact across F1's MarTech operations:

- Data source onboarding: reduced from 6 to 8 weeks to approximately 40 minutes of code generation plus hours of deployment and review.
- Autonomous work: AI agents handle 95% of onboarding tasks without human intervention.
- Time-to-value: approximately 99% reduction.
- Schema evolution: end-to-end resolution in hours instead of days.
- Integration backlog: 18-month backlog cleared in weeks.
- Data engineers who previously spent their time writing boilerplate ingestion code and chasing schema breaks now focus on strategic initiatives that advance the business.
- Implementation velocity: a single developer took the agentic solution from proof of concept to production release in 4 months.

The reliability, consistency, and data integrity of the MarTech platform were improved, while the operational overhead was reduced: "The Data Accelerator didn't just speed things up. It changed how we operate. Our data engineers went from writing boilerplate ingestion code to focusing on strategic initiatives. Issues can be identified and fixed before our end users even notice." says Kemp.

Conclusion

The Data Accelerator's success comes down to three principles: meeting developers where they already work, keeping them at the helm, and embedding governance like GDPR classification from day one rather than bolting it on after. These principles shaped a solution where F1 partnered with AWS to use agentic AI on Amazon Bedrock AgentCore to transform MarTech data operations. By combining automated data source onboarding, schema evolution detection, and unified data access through Amazon SageMaker Unified Studio, F1 reduced onboarding time by approximately 99% and eliminated an 18-month integration backlog in weeks.

The approach is deliberately replicable. Any organization dealing with multi-source data onboarding, schema volatility, and governance requirements can apply the same architecture to their own environment. The agents are domain-agnostic, and they know how to onboard, classify, and monitor. The domain is interchangeable.

Getting started

To learn more about the AWS services used in this solution:

- Amazon Bedrock AgentCore.
- Amazon SageMaker Unified Studio.
- Kiro.
- AWS Professional Services.

Acknowledgments

This outcome is the result of years of incremental improvements to the MarTech platform, delivered through a close partnership between F1 and AWS. Many contributors across both organizations have shaped the architecture and strengthened the foundations that made the Data Accelerator possible. We are grateful to the following thought leaders and developers for their dedication and expertise: Paula Marenco Aguilar, Nadeen Nilanka, Taye Aduewa, Marton Juhasz, Deepak Gulia, Alex Goff, Nick Morgan, and Seshadri Senthamaraikannan.

About the authors

Subhro Bose

Subhro is a Senior Data & AI Architect at AWS and the creator of CausalIF, an open-source causal inference library that brings causal inference to model reasoning with consistency of results, discovering why things happen in complex systems. His work spans agentic AI, self-healing platforms, and turning causal discovery into production-grade intelligence across logistics, finance, and compliance.

Jerome Descreux

Jerome is a Senior Delivery Manager within AWS Professional Services. He leads large scale transformation programs and strategic projects for major EMEA enterprise customers, across various industries including logistics, manufacturing, financial services, aviation, and sports.

Matt Kemp

Matt leads CRM and Customer Data Operations at Formula 1, overseeing the data, insight and engagement capabilities that connect millions of fans with the sport. He has delivered large-scale digital transformation programmes and is championing the use of AI, machine learning and agentic technologies to drive innovation, operational efficiency and personalised fan experiences.

Arunraja Kumar

Arunraja is Senior Data Architect at Formula 1, responsible for shaping the data platform and architecture that powers fan engagement, insight and innovation across the sport. He led the technical transformation of F1's Fan Personalisation Programme, evolving the platform into an AI-native, agentic ecosystem that enables scalable, real-time experiences and supports Formula 1's ambition to deliver personalised engagement to more than one billion fans worldwide.