

Reducing Errors and Building Trust with the DataKitchen Platform

Data errors in production are a significant issue for most organizations. In a recent [DataOps survey](#), eighty percent of companies surveyed reported three or more errors per month. Thirty percent of respondents reported more than 11 errors per month. However, organizations that follow DataOps principles typically have less than one error per year. The experience of a large transportation company illustrates how following DataOps principles can eliminate costly and embarrassing errors and transform the data team into a trusted partner.

THE PROBLEM

A large transportation company experienced significant numbers of data errors, and consequently, a lack of trust with its business users. Although the company strove to be data-driven, a complicated organizational structure and data infrastructure made it extremely difficult to identify and manage errors as data flowed through the end-to-end system.

A tremendous amount of data was collected from different sources, including vehicles and other internal systems. Like many companies, they had several data architectures – batch, streaming, big data, small data, on-premises, cloud, and prescriptive and predictive models – all working together (Figure 1).

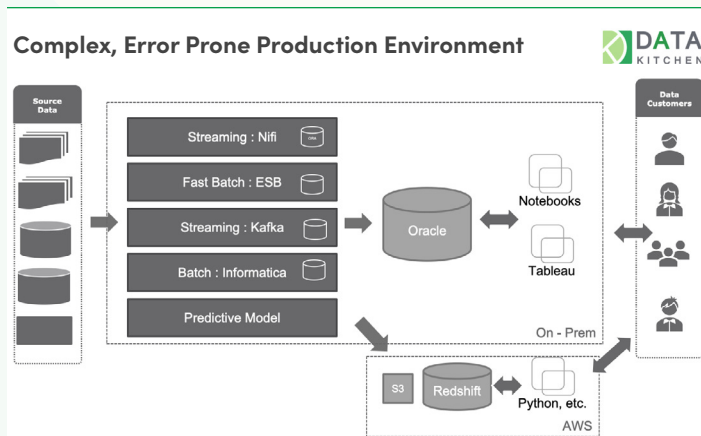


FIGURE 1:
A complicated data infrastructure made it extremely difficult to identify and manage errors in data analytics.

In addition, different teams managed the creation and operation of these pipelines in different locations (Figure 2). As a result of this complexity, errors often went

Complex Team Development Ownership

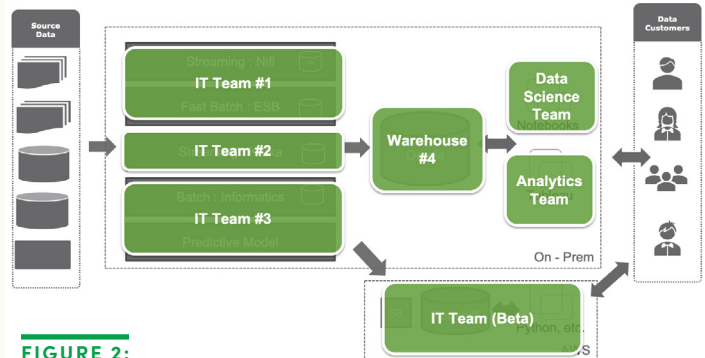


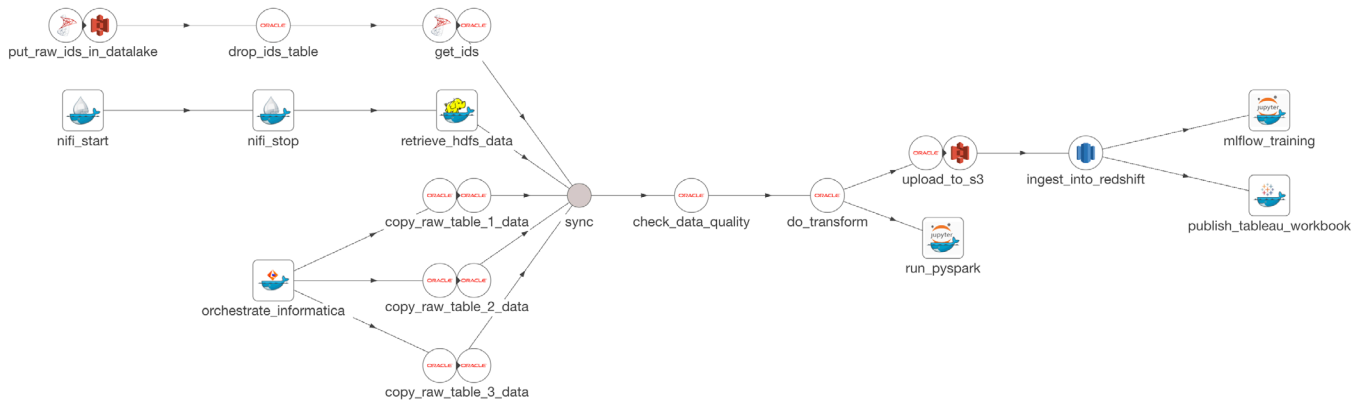
FIGURE 2:
Different teams managed the creation and operation of data pipelines in different locations.

undiscovered until they reached the end-user. This situation led to chaos, stress, and wasted time as the data team often scrambled to identify and fix errors.

THE SOLUTION

To reduce errors, the team implemented DataOps using the DataKitchen DataOps Platform. DataOps is a set of technical practices, cultural norms, and architectures that enables, among other benefits, enhanced collaboration and low error rates. The DataKitchen Platform supports DataOps through automation of the end-to-end data analytics factory, from data access to value delivery.

The data team decided to first focus on automated orchestration and monitoring of their multi-technology pipelines. The team was not interested in changing any of their existing tools, and the DataKitchen tool-agnostic platform gave them the ability to seamlessly orchestrate a heterogeneous toolchain. The 'Recipe' in Figure 3 automates the orchestration of the team's diverse range of tools and processes.

Reducing Errors and Building Trust with the DataKitchen Platform *(Continued)***FIGURE 3:**

The graph displays DataKitchen's automated orchestration and monitoring of the data team's multi-technology pipelines.

In addition, automated tests were embedded in every step of the pipeline to monitor quality and catch errors before they reached the next stage. These tests run continuously in production. Using the Platform, the team was able to write tests in their tools of choice. They incorporated a wide range of tests, such as basic data quality tests, statistical process control, location balance, historical balance, and business heuristic tests.

Notifications were configured so that an alert would be sent via Jira, email, or Slack to the appropriate team members when an issue was detected. As a result, the team would be able to catch and address any error before it moved down the pipeline, improving the quality and speed of the data moving through the system.

RESULTS

The implementation of DataOps quality testing and automated orchestration demonstrates how the DataKitchen Platform is able to transform data-analytics organization quality, culture, and workflows.

The new DataOps process dramatically reduces errors. The elimination of errors transforms a culture of shame and blame into one of trust. As a result, the team can expect to spend less time putting out fires, and more time creating new analytics that deliver business value to their customers.

Without changing any of their existing tools, automated orchestration enables the team to quickly and seamlessly transform data into business insight. The 'single pane of glass' view of the entire end-to-end analytic process reduces complexity and improves collaboration and transparency across the organization. The pipeline visualization enables the entire team to understand what their colleagues are working on and how their piece fits into the bigger picture.

The next step for the team is to extend their DataOps capabilities to include automated deployment of new analytics. Combining automated deployment with orchestration and testing, will lead to even greater improvements in analytic cycle time.