CASE STUDY

Accelerating Analytic Cycle Time with DataOps and the DataKitchen Platform

datakitchen.io info@datakitchen.io

DataKitchen

In today's on-demand economy, the ability to quickly derive business value from data will separate the winners from the losers. Yet, many organizations experience excruciatingly slow cycle times for creating analytic environments or deploying new analytics. According to an Eckerson survey, 76% of organizations are too slow to deploy, facing deployment times of weeks, or even months. This is typically due to manual processes with little to no automation or testing. Data organizations that neglect to modernize their processes, risk being left behind. However, the good news is that any company can dramatically improve their cycle time with a DataOps approach to analytics.

THE PROBLEM

A large telecom company was experiencing **cycle times of greater than four months** to release new features into production, impeding their ability to provide the on-demand analytic insights needed to fuel their business. Although the team was extremely capable, their established processes emphasized quality over speed and agility.

The process involved a four-stage manual deployment from development to production (see Figure 1). Like most large enterprise organizations, the pipeline involved the coordination of a wide range of tools, environments, and teams, which introduced complexity, slowness, and errors.

Furthermore, the process did not include any automated testing. Any testing that did take place was manual with no coordination or feedback loops between environments or teams. Whenever an error was detected, it kicked off a lengthy process that involved many steps – completing a bug report, notifying developers, attempting to reproduce the error, fixing the problem, and running QA again. If the error persisted, the steps were repeated again, adding significant time to the process.

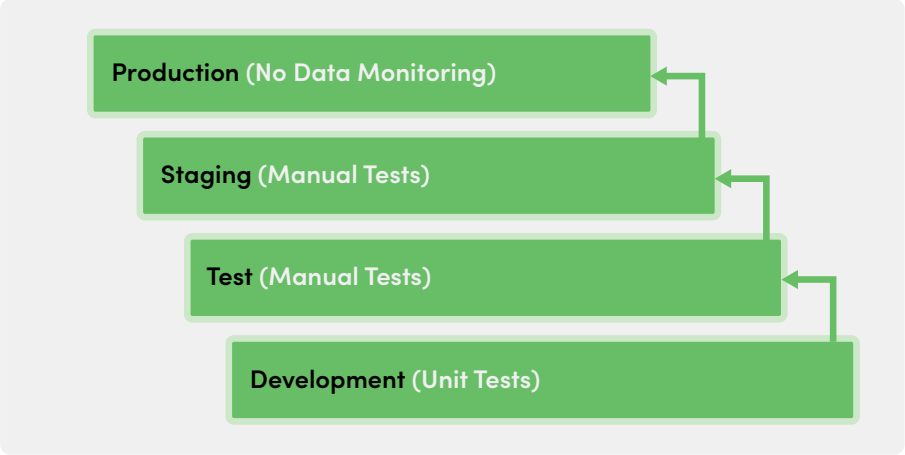

FIGURE 1: *Analytic cycle time involved a 4-stage manual deployment from development to production.*

Worse still, many errors were slipping through to production. These errors were often caught by the business users, reducing the overall level of trust in the data analytics organization.

THE SOLUTION

To improve cycle time, the team implemented DataOps using the DataKitchen DataOps Platform. The team started small and identified a bottleneck in its core EDW pipeline, that if alleviated could yield a significant business benefit. Within this pipeline, they focused their initial efforts on the first two steps (Development and Test) (see Figure 1).

As a first step, the team created separate but aligned environments. The DataKitchen Platform enabled the team to easily to spin-up development environments that were clones of the production environment. The more similar the environments, the easier it is to test, deploy code to production and, if needed, replicate production errors.

The team also introduced automated testing. The use of automated tests reduces cycle time and decreases the time required to find and fix issues, thus speeding new

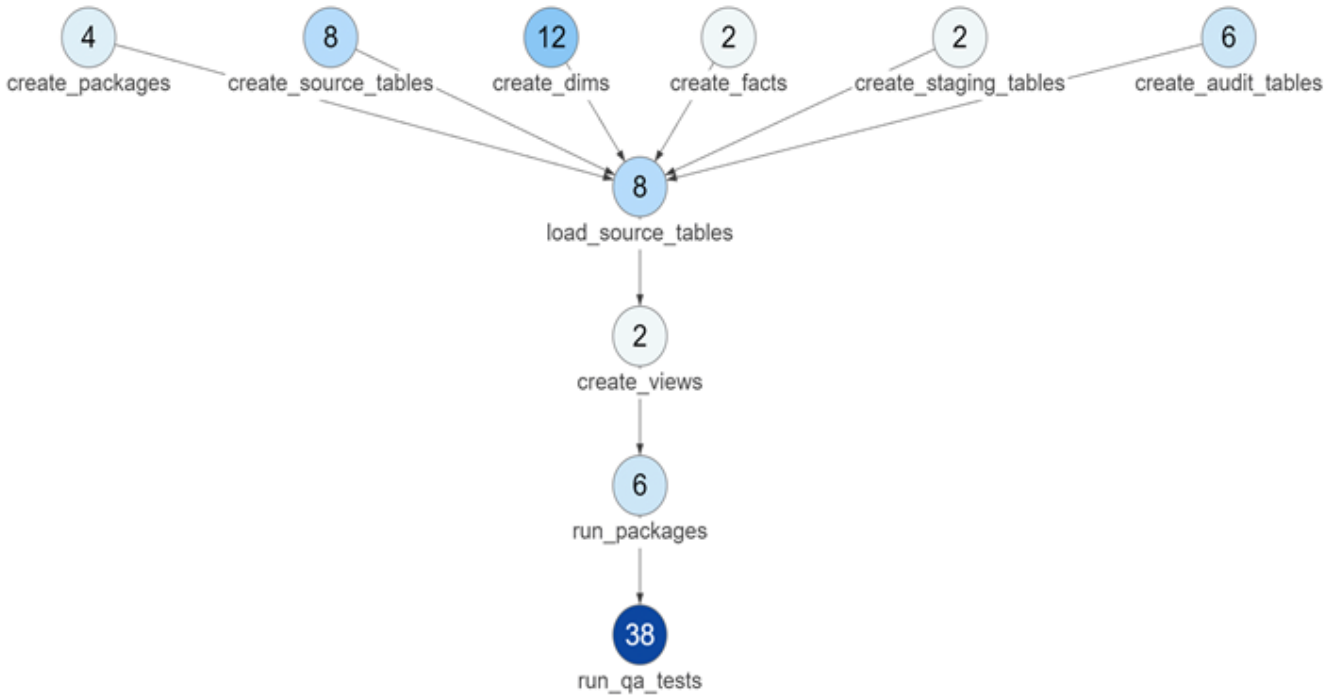

FIGURE 2: *Using the DataKitchen Platform, the team embedded multiple automated tests at every step in the pipeline.*

features into production.

Using the DataKitchen platform, the team embedded automated tests at every step in the pipeline to make sure that when analytics moved from a development environment through each of the other separate environments, that everything would still work (see Figure 2). Errors are now identified well before they reach production and the business users.

In addition to standard unit tests, the team included a wide breadth of tests like functional, regression, performance tests, to name a few. A location balance test was used to ensure no loss in rows as data progressed through each environment. Because of the Platform's standardized interface, each test could be written in the user's tool of choice.

Test result histories are also now available and easy to understand by other team members, even if they did not create the tests. When issues arise, users can instantly identify and address the root cause of errors, and add new tests to increase the robustness of their pipeline over time.

RESULTS

DataOps results in dramatic improvements in the time required to create new analytics. Developers can now catch errors in real-time and fix them immediately, avoiding the lengthy multi-step process that was previously used to identify and fix issues. Less time spent fixing errors also means that the team can spend more time on innovation.

For even greater time savings, the next step for the team is to automate the deployment of new analytics into production. They can also add automated DataOps monitoring to their production pipeline to ensure that all the data is running smoothly and efficiently through the pipeline without errors. By implementing these core DataOps principles, the team will be well-positioned to deliver the trusted, high-quality, on-demand insight that is critical to drive the business forward.