

Improving Flaky Tests in Data Test Automation Pipelines

Dhanunjay Reddy Seelam

Senior Software Engineer, Bentonville, United States

Abstract: *Flaky tests in data automation pipelines can cause unreliable test results, low developer confidence, and inefficient development cycles. These tests typically have a high degree of inconsistency in terms of pass/fail outcome, a phenomenon often due to unstable environments, changing data, or test logic that is bad to begin with. The paper thoroughly investigates test flakiness in data pipelines, offering detailed insights into its effects on development and quality assurance (QA) activities. The document delves into strategies for mitigating flakiness, including stabilization techniques for data, better test design, and infrastructure improvements. Other methods like root cause analysis powered by AI and self-healing automation are also considered to make these tests stable and reliable. By leveraging a combination of case studies and concrete empirical evaluations, the efficacy of these various strategies is shown, providing real-world guidance as well as frameworks for practitioners looking to develop more resilient and dependable data test automation pipelines. The findings of this research add to the growing body of work aimed at improving software testing methodologies, highlighting the importance of having stable and deterministic test results in our increasingly data-centric applications.*

Keywords: Flaky tests, Data Pipelines, Quality Assurance, Artificial Intelligence, Test Automation, Software Testing