WHEN A GREEN PIPELINE IS NOT ENOUGH

Applying STPA to CI/CD Database Validation Decisions

The validation tool can be correct.

The pipeline can be green.

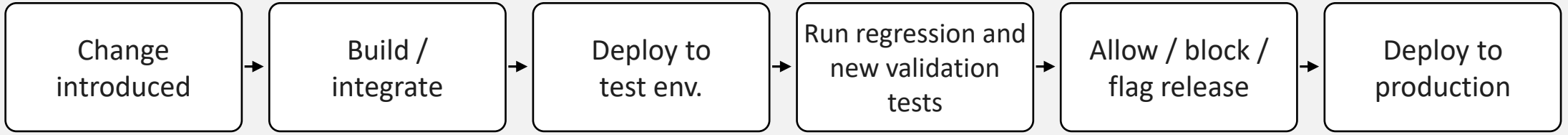
The decision can still be wrong.

Gil Shaul-Shay, Dr. Avi Shaked | Tel Aviv University

TEL AVIV UNIVERSITY אוניברסיטת תל אביב

A CI/CD DECISION

Continuous Integration and Continuous Delivery (CI/CD) is an automated workflow that moves a change through build, test, validation, and release decisions.



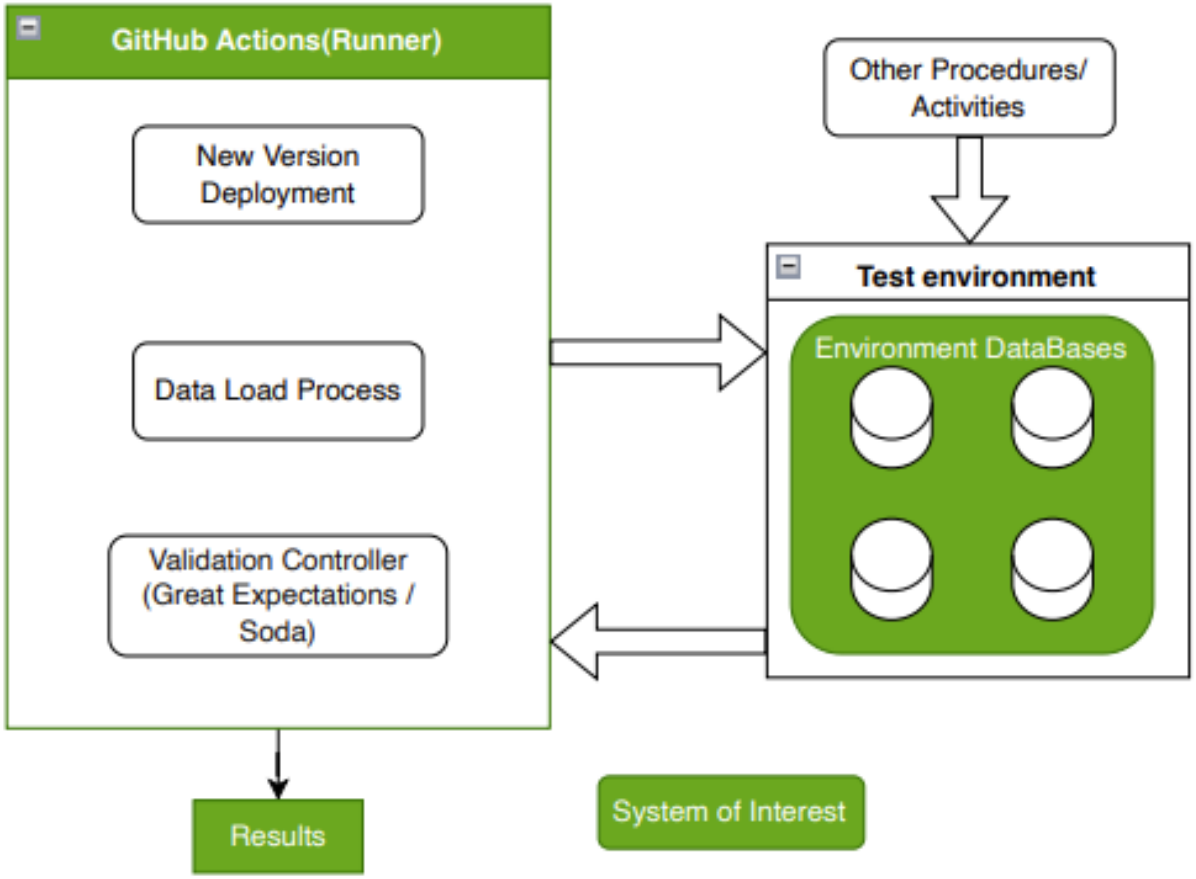
Why this becomes hard for testers, QA, and integration engineers

The software keeps changing: features, configuration, data, and environments.

Some validation and regression tests are reused from earlier versions and assumptions.

Symptoms: flaky results, false failures, and false passes.

REPRESENTATIVE CASE STUDY AND BOUNDARY



WHAT DOES A “GREEN” TEST RESULT ACTUALLY PROVE?

A green result tells us

- The executed tests passed.
- The validation tool completed its run.

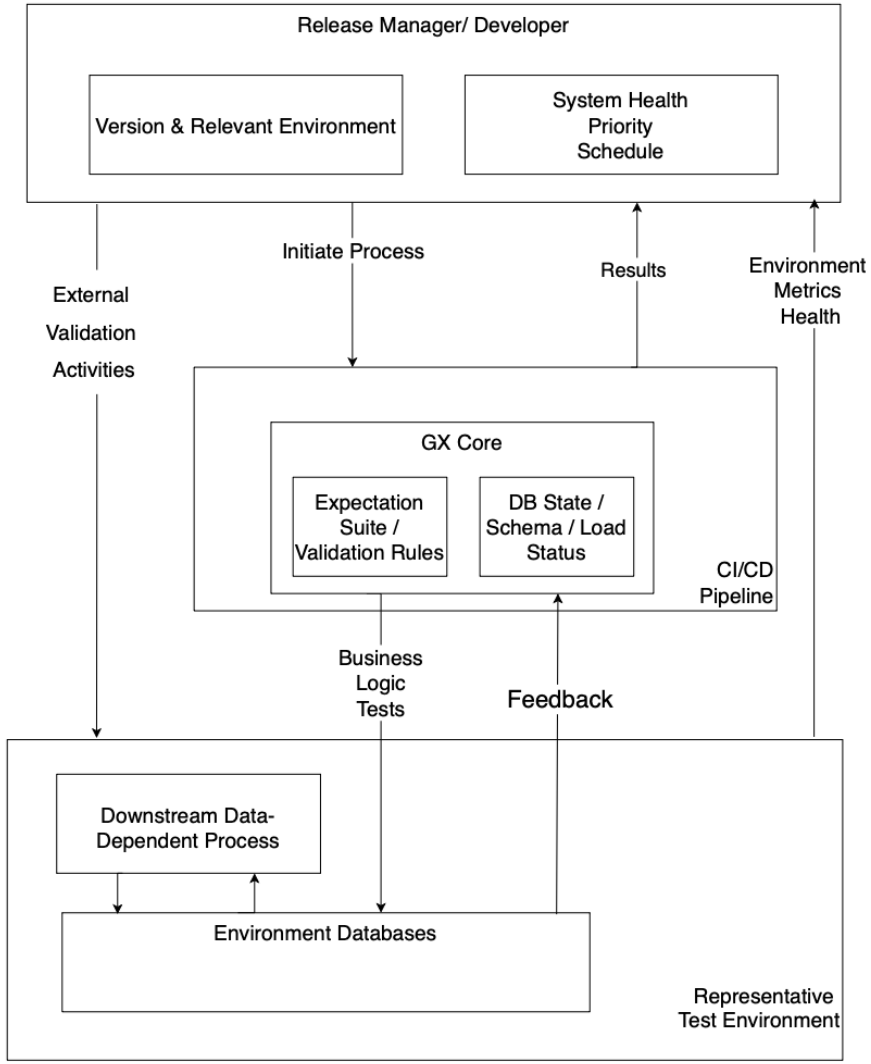
It may not tell us

- The data load was complete and fresh.
- The database state did not change during validation.
- The validation rules still match the schema and assumptions.
- The environment was ready and representative.
- The result is still applicable at decision time.

A green result is not enough if it does not represent the database state used for the decision

THE CONTROL STRUCTURE CHANGED THE QUESTION

This is the shift from “Did the tests pass?” to “Can we rely on this evidence for the specific allow / block / flag decision?”



WHAT DID THE ANALYSIS IDENTIFY?

System-level losses

- **L-1** Loss of integrity or trustworthiness of downstream data-dependent operation.
- **L-2** Loss of operational continuity: required data/version updates are delayed, blocked, repeatedly rerun, or used with unreliable evidence.
- **L-3** Loss of confidence in automated validation because results are flaky, unexplained, or misleading.
- **L-4** Loss of downstream service or decision quality due to invalid or incorrectly validated data.

System-level hazards

- **H-1** Invalid, incomplete, stale, duplicated, corrupted, or inconsistent data is made available downstream as trusted data.
- **H-2** The pipeline assigns trusted validation status to a data/version state that does not represent the actual database state downstream processes will use.
- **H-3** A promoted data/version state violates the current schema, validation assumptions, or downstream data expectations while being considered acceptable.
- **H-4** A required update remains unavailable because the pipeline cannot produce a timely trusted, blocked, or flagged decision.
- **H-5** The pipeline treats a data/version state as usable when its validation status is unknown, uncertain, expired, or invalidated by later changes.

ONE CONNECTED PATH THROUGH THE ANALYSIS

L-1
Loss of integrity or trustworthiness of downstream data-dependent operation.

H-2
The pipeline assigns trusted validation status to a data/version state that does not represent the actual database state downstream processes will use.

SC-2
Validation status shall be trusted only when the validated database state represents the state intended for downstream use or version promotion.

UCA-2
The validation controller executes validation tests while the database state being validated is modified by another workflow, background job, or manual update.

LS-2
The database state changes during the validation window, but the pipeline does not detect that the validated state no longer matches the state used for the decision.

Design change

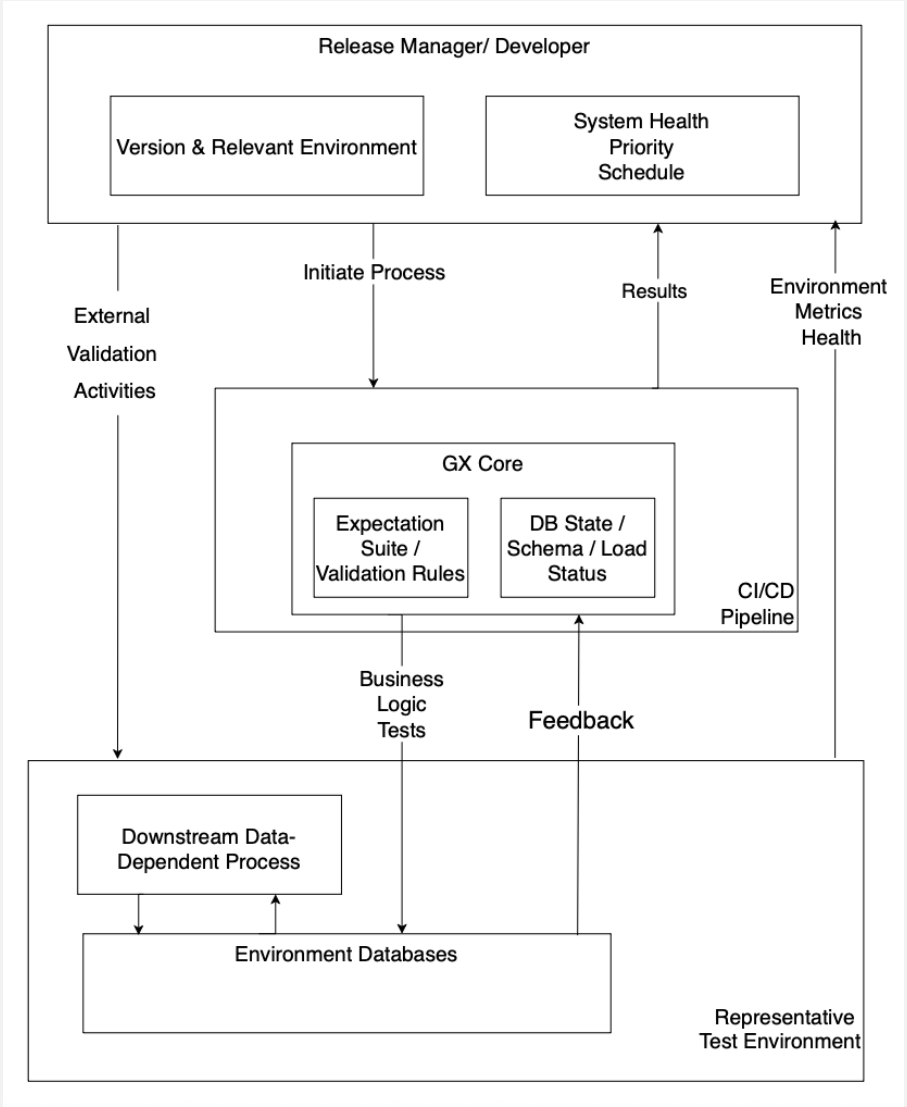
Monitor database state and execution conditions during validation. If state, configuration, activity, or health changes, discard, flag, or revalidate the result.

STPA-DRIVEN ARCHITECTURAL REDESIGN

Context & Consistency
CI/CD metadata integration
Configuration consistency validation

Environment Awareness
Environment activity detection
Real-time monitoring integration
Readiness verification

Result Integrity & Enforcement
Validation result integrity verification
Supervisory decision logic



PRACTICAL VALUE, EFFORT, AND NEXT STEPS

Value demonstrated so far

STPA reframed flaky / misleading validation as feedback and process-model gaps, not only test-script defects.

The analysis produced concrete design changes around test evidence: completion, freshness, readiness, compatibility, integrity, and revalidation.

The contribution is a publicly shareable STPA application in CI/CD validation and non-safety quality/reliability decisions.

Effort to Apply STPA

- 1 primary analyst, working part-time as M.Sc. research
- Advisor review and iteration on the STPA artifacts
- STPA experience: learned/applied during the research
- Focused STPA effort: 45 hours
  Substantial review iterations / meetings: 5

Prototype

Two of the three planned prototype parts have been completed.
Controlled representative scenarios derived from the STPA loss scenarios compare: baseline pass/fail behavior vs. STPA-supervised behavior.

Questions?

gilshaulshay@mail.tau.ac.il