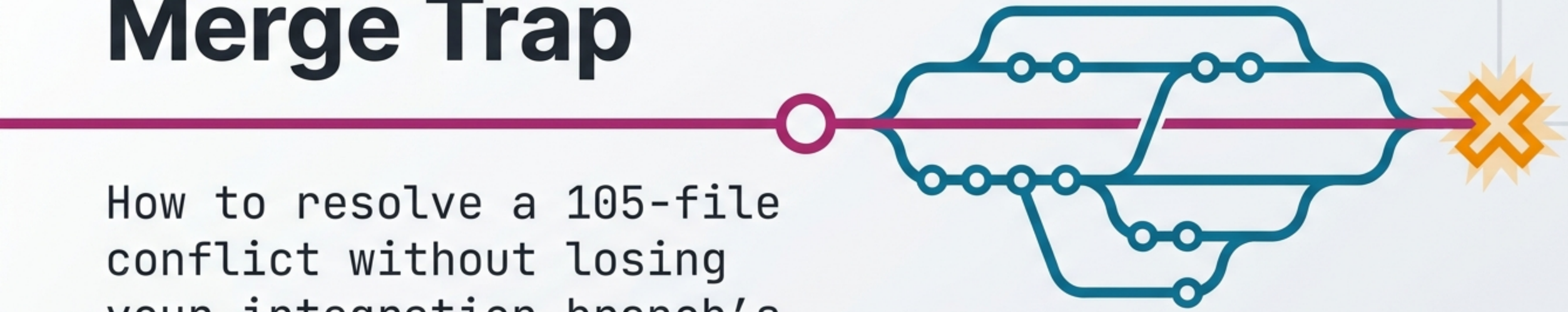
The Squash-vs-Granular Merge Trap



How to resolve a 105-file conflict without losing your integration branch’s history.

NotebookLM

Trunk (main)

- SIEM log-lake cutover landed 2 months ago.
- One squash commit. SHA `14a0661`.
- 224 files touched, +35,469 / -7,267 lines.

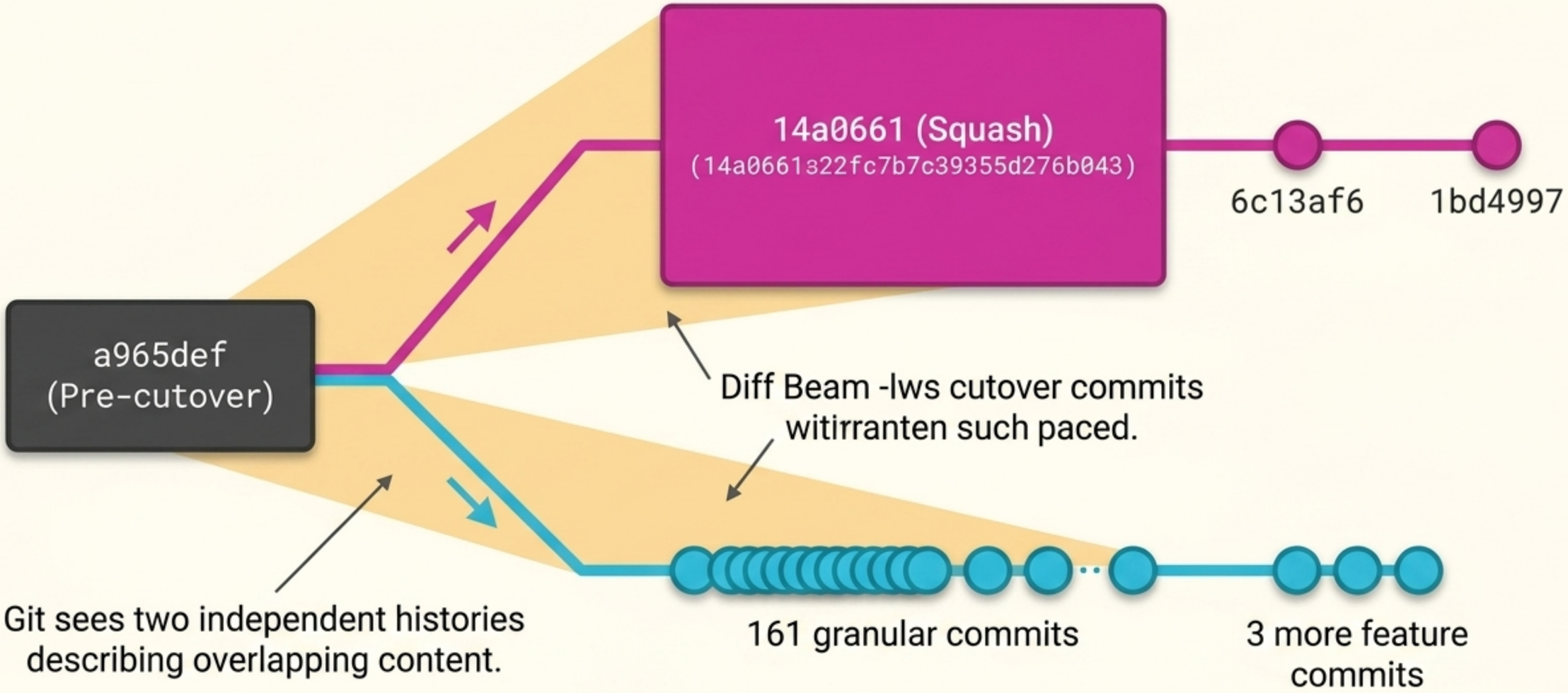
Integration Branch (siem-loglake)

- Built on the same cutover base.
- 161 TDD-shaped granular commits.
- 13 additional feature commits on top.

```
$ git merge origin/main
 CONFLICT (content): Merge conflict in 105 files.
```

Plain `git merge` could not do it. 174 commits ahead of an ancestor that no longer existed on `main`.

NotebookLM

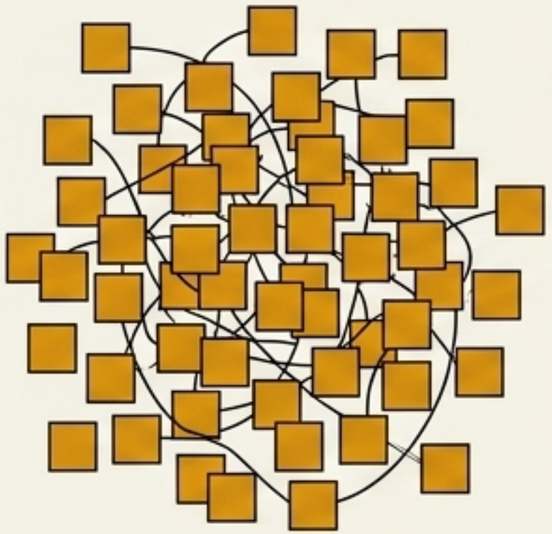


The Trap: Git compares both tips against `a965def`. Because the squash destroyed per-commit lineage, identical content registers as a massive two-sided conflict.

NotebookLM

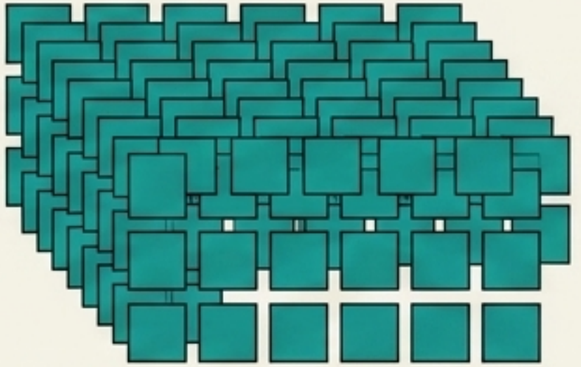
Conflict Triage: Categorization Beats Resolution

105 Raw Conflict Files



Categorization Beats Resolution

94 Pure Cutovers
Mechanically Resolvable



11 The Careful Set
Per-File Decision Required

Most of the 105 conflicts are conflicts in name only. Taking the integration branch’s evolved version is correct for the vast majority of files.

NotebookLM

```
$ git diff --name-only <squash-commit> origin/main
  git diff --name-only <squash-commit> origin/main

queries.py
routers/dns.py
HealthStrip.tsx
... (10 files total)
```

Give me just the paths.

Show me what is genuinely new on `main` since the squash landed.

Summary

This command isolates the 10 files that `main` actually evolved. Out of 105 files, everything else is pure cutover. You now know exactly where to spend your manual effort.

NotebookLM

Diagnostic Matrix: Conflict Categorization & Resolution Strategy

Group	Count	Profile	Resolution Strategy
Pure cutover & CI-reformat	94	Unchanged by trunk since squash.	`git checkout --ours`
Conflicted Supersets	5	Trunk added net-new features (e.g., DNS search).	Take `--theirs` after eyeball verification.
Small / Blank-line Conflicts	3	Minor drift in imports or spacing.	Manual in-place edit.
Bidirectional (queries.py)	1	Both branches added unique, non-overlapping logic.	Base on `--ours`, awk/sed splice in main's block.
Isolated Mock Handlers	1	Branch has unique SIEM mocks.	Take `--ours`.

NotebookLM

Treatment 1: The Mechnical Loop

Treatment 1: The Mechanical Loop

```
cat /tmp/group1.txt | while read f; do
  git checkout --ours "$f"
  git add "$f"
done
```

The list of 94 pure cutover files.

Counterintuitive Merge Convention:

When you run `git merge origin/main` from your reconcile branch, `--ours` means your starting side (siem-loglake). `--theirs` means main.

The rule: `--ours` is whoever's tip your `HEAD` was at when you started the merge.

Time to resolve 94 files: **< 1 minute**.

NotebookLM

Treatment 2: Superset Eyeballing

:2:routers/dns.py (Ours)

```python
import logging
from fastapi import APIRouter, Depends
from services.dns_resolver import resolve_host

router = APIRouter()
logger = logging.getLogger(__name__)

# Basic setup complete.
```

:3:routers/dns.py (Theirs)

```python
import logging
from fastapi import APIRouter, Depends
from services.dns_resolver import resolve_host

router = APIRouter()
logger = logging.getLogger(__name__)

# Basic setup complete.

@router.get('/api/dns/search')
async def search_dns(query: str):
    logger.info(f"Searching DNS for: {query}")
    return resolve_host(query)
```

Command:
`git diff :2:<file> :3:<file>`

→

Verification: Is ‘theirs’ a clean superset of ‘ours’?

→

Action: If yes ->
`git checkout --theirs <file>`

Files like `test_migration_0006.py` and `HealthStrip.tsx` were strictly more functional on main. Verifying supersets allows safe, wholesale adoption of the trunk’s version.

NotebookLM

Treatment 3: The Bidirectional Splice

The Target: `:3:backend/src/.../queries.py`

- Integration branch fixed `argMax`. Main added `dns_search`. Both must survive.

Reads the conflicted theirs-side directly from the index without touching working tree.

```
git show :3:backend/src/.../queries.py | \
  sed -n '482,652p' | \
  awk '/def dns_search/ {print; next} {print}' > splice.tmp
```

Sitracts the block known good contiguous block from **main**.

Extracts the exact known good contiguous block from **main**.

Inserts the block exactly at the target anchor pattern.

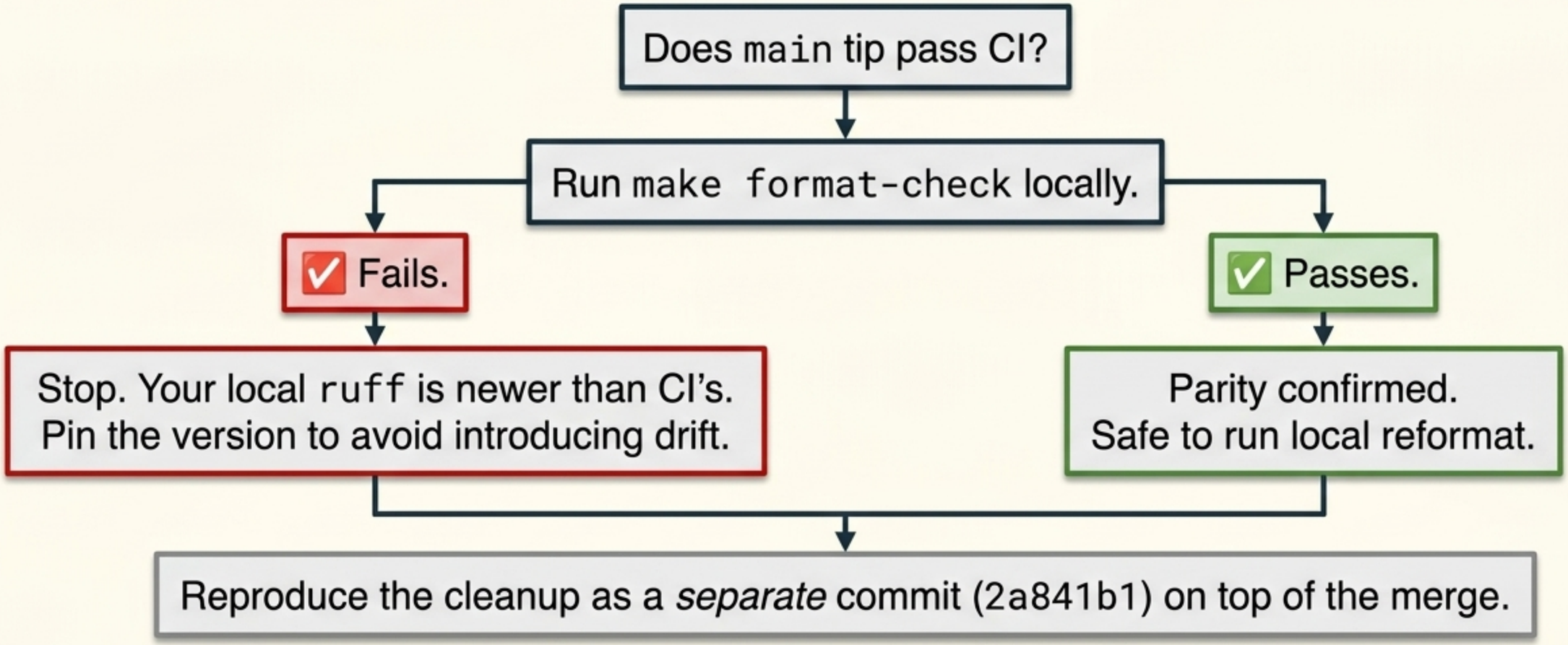
✅ **Verification Check:**

- ✅ `py_compile` (confirms valid syntax)
- ✅ `grep` (confirms both `argMax` and `dns_search` exist in output file)

NotebookLM

Complication 1: The Skipped Trunk Cleanup

`Main` **applied** `ruff --fix` **after the squash**. When you took `--ours` on the 94 files, you accidentally pulled the un-reformatted legacy versions.



NotebookLM

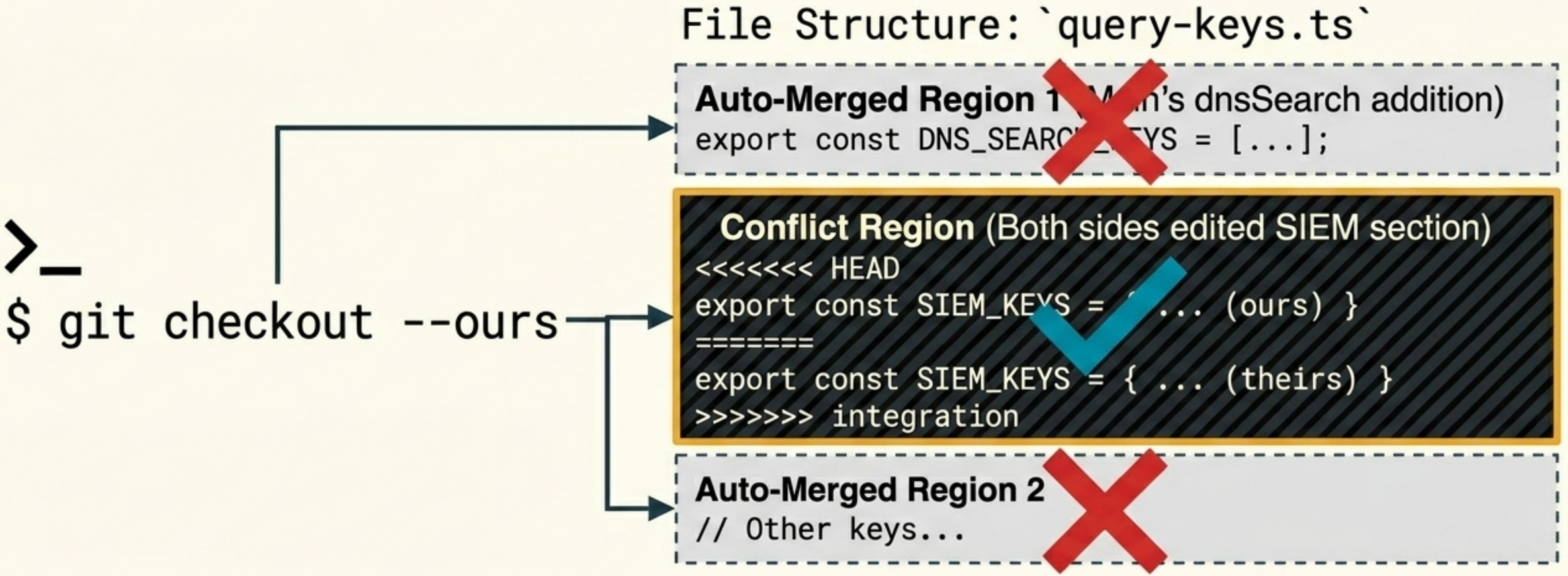
Complication 2: The zip(strict=True) Autofix Trap (B905)

The Silent Contract	The Autofix Behavior Change
`zip([1,2,3], [A,B])`	`zip([1,2,3], [A,B], strict=True)`
Result: **[(1,A), (2,B)]** (Silently drops the extra element)	Result: **ValueError**
Existing codebase relies on tolerant “zip to shortest” contract in production polling registries.	Ruff autofixes all call sites. Breaks test_writer_called_with_merged_findings.

Do not blindly accept autofixes that change runtime behavior.
Production polling registries got explicit ‘strict=False’ to preserve the contract.

NotebookLM

The Danger of `git checkout --ours`



WARNING: `git checkout --ours` takes the *whole* ours version. It silently drops any auto-merged content from their side in non-conflicted regions. **Do not** use this command for bidirectional edits.

NotebookLM

In-Place Editing Heuristics

When to use `--ours` or `--theirs`

- The file is pure cutover, or one side is a strict superset of the other.

`Fast, whole-file overwrite.`

When to Edit In-Place

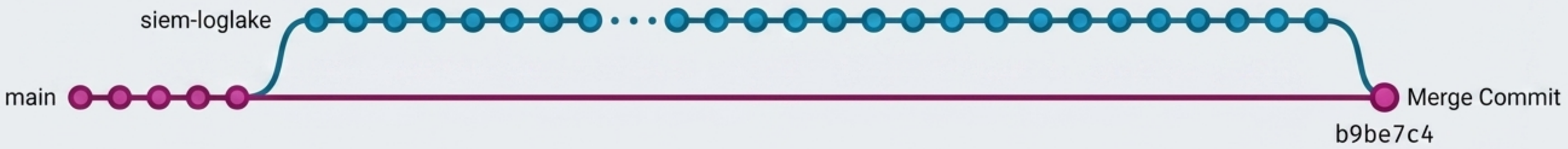
- File is in the "Careful Set" (main has unique content).
- `git diff` shows multiple hunks per side, or short conflicted hunks.

`Auto-merged regions are preserved automatically.`

Defensive Default: If a file is in the careful set, do not run checkout commands without reading `git diff :2: <file> :3: <file>` first.

NotebookLM

PR #23 merged. **+12,118 / -32. Purely additive.**



Close the Loop

```
git checkout siem-loglake
git push origin main:siem-loglake
```

Fast-forwarding the integration branch ensures it is a strict ancestor of main. There is no divergence left to recur. Any new work starts from a clean base.

NotebookLM

The Squash Reintegration Playbook

- [x] 1. Start a merge on a throwaway reconcile branch (`git merge origin/main`). Do not commit.
- [x] 2. Categorize conflicts with `git diff --name-only <squash-commit> origin/main` to find the trunk's 'genuinely new' files.
- [x] 3. Mechanically resolve pure cutover files (`git checkout --ours` via `while` loop).
- [x] 4. Hand-merge the 'careful set' using `git diff :2:<f> :3:<f>` to verify supersets, and in-place edits to preserve auto-merges.
- [x] 5. Reproduce trunk tool cleanups (e.g., `make format`) as a separate commit. Verify local vs CI parity first.
- [x] 6. Fast-forward the integration branch after the merge lands to prevent recurrence.

`4 hours clock time. 174 granular commits preserved. 0 history lost.`

NotebookLM