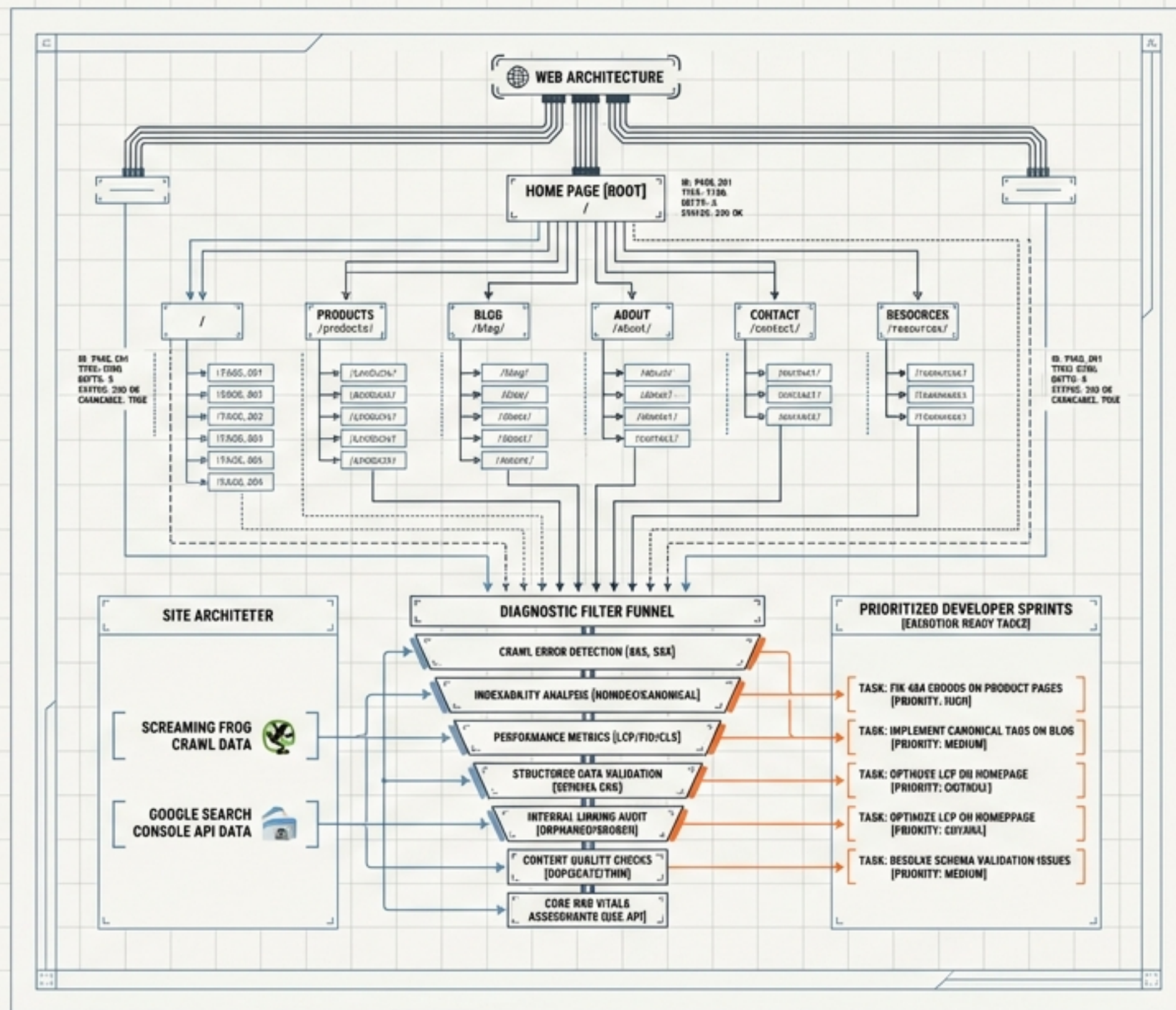


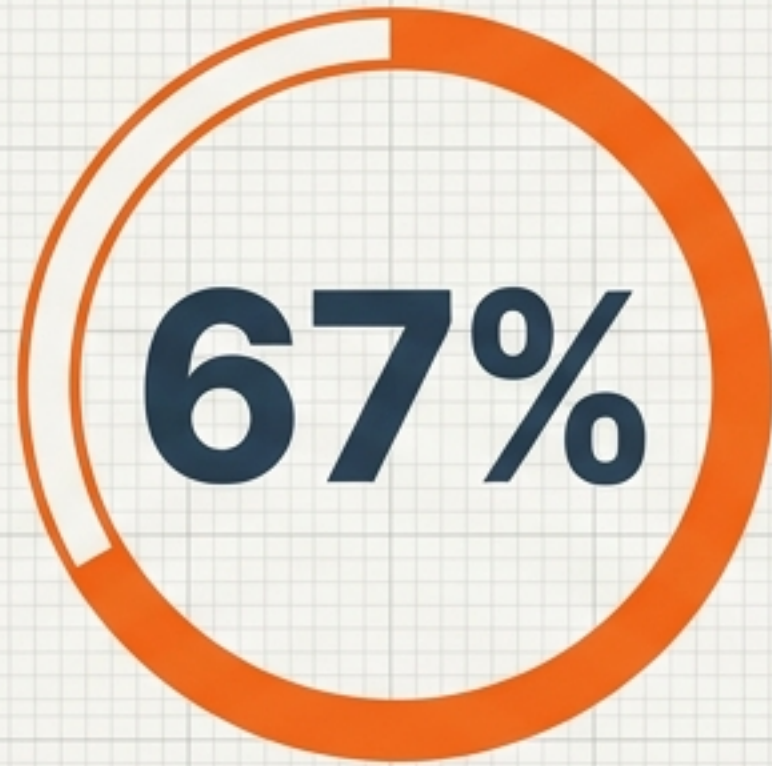
Raw Data to Developer Sprints: The Technical SEO Audit Playbook

A diagnostic framework for translating Screaming Frog & Google Search Console data into prioritized, execution-ready tasks.



[SYS_REQ: SCREAMING_FROG_v24.1 // GSC_API_CONNECTED]

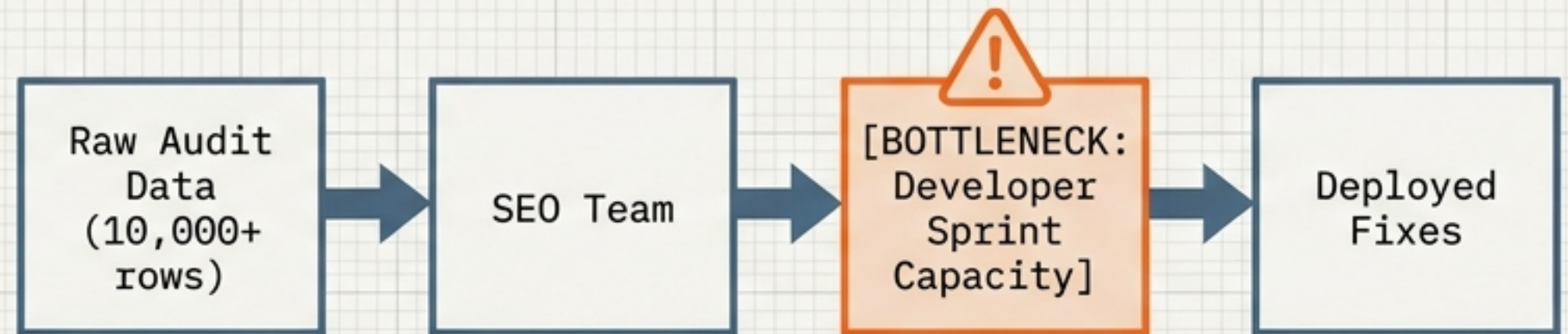
The Reality Check: 67% of teams are blocked by developer bandwidth.



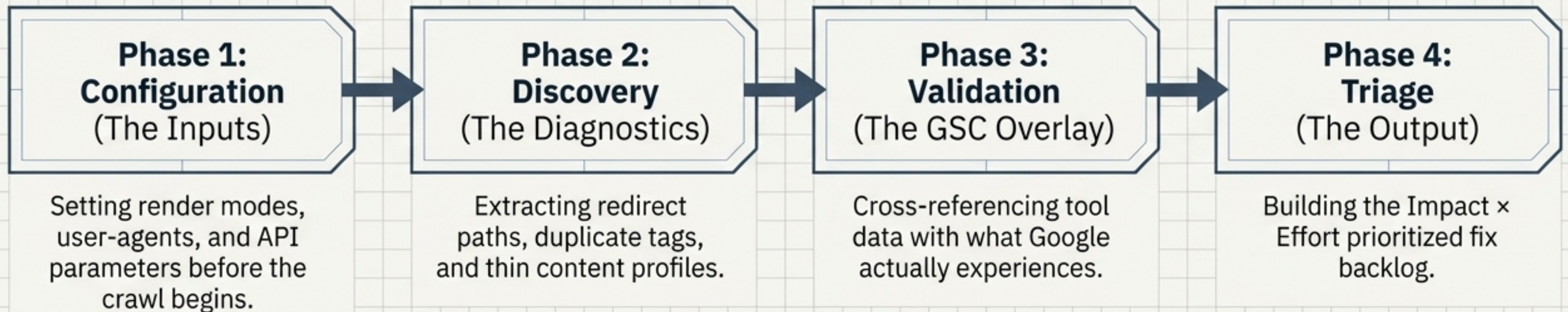
The Problem: Finding technical errors is easy; getting them fixed is the barrier.

The Solution: Priority order is the only lever the SEO team controls.

The Mandate: We must diagnose, score, and triage organic issues with surgical precision.



The Diagnostic Blueprint: A four-phase system



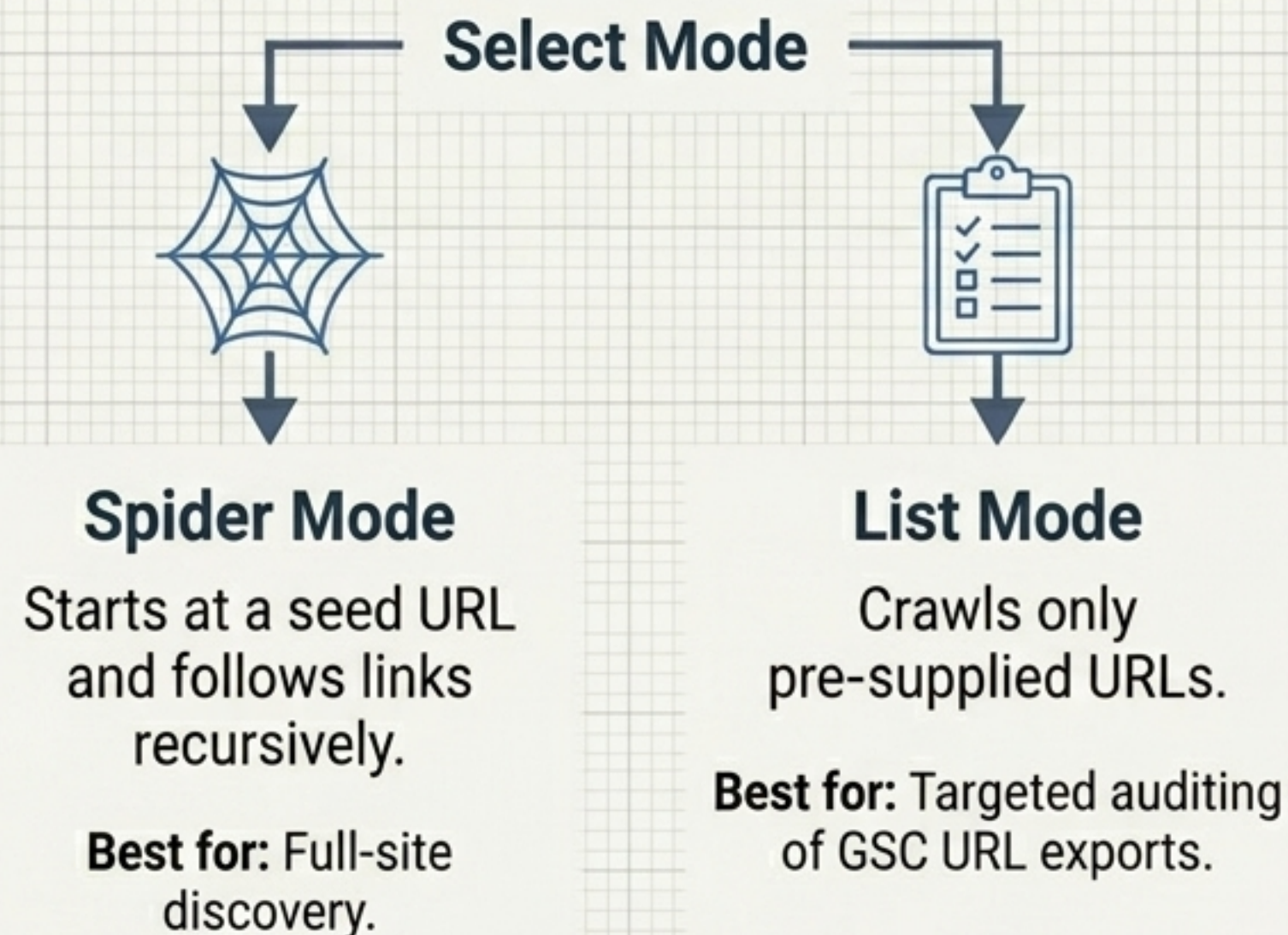
OUTPUT_GOAL: Developer-Ready Sprint Tickets

Step 1: Establish your operational boundaries and crawl mode.

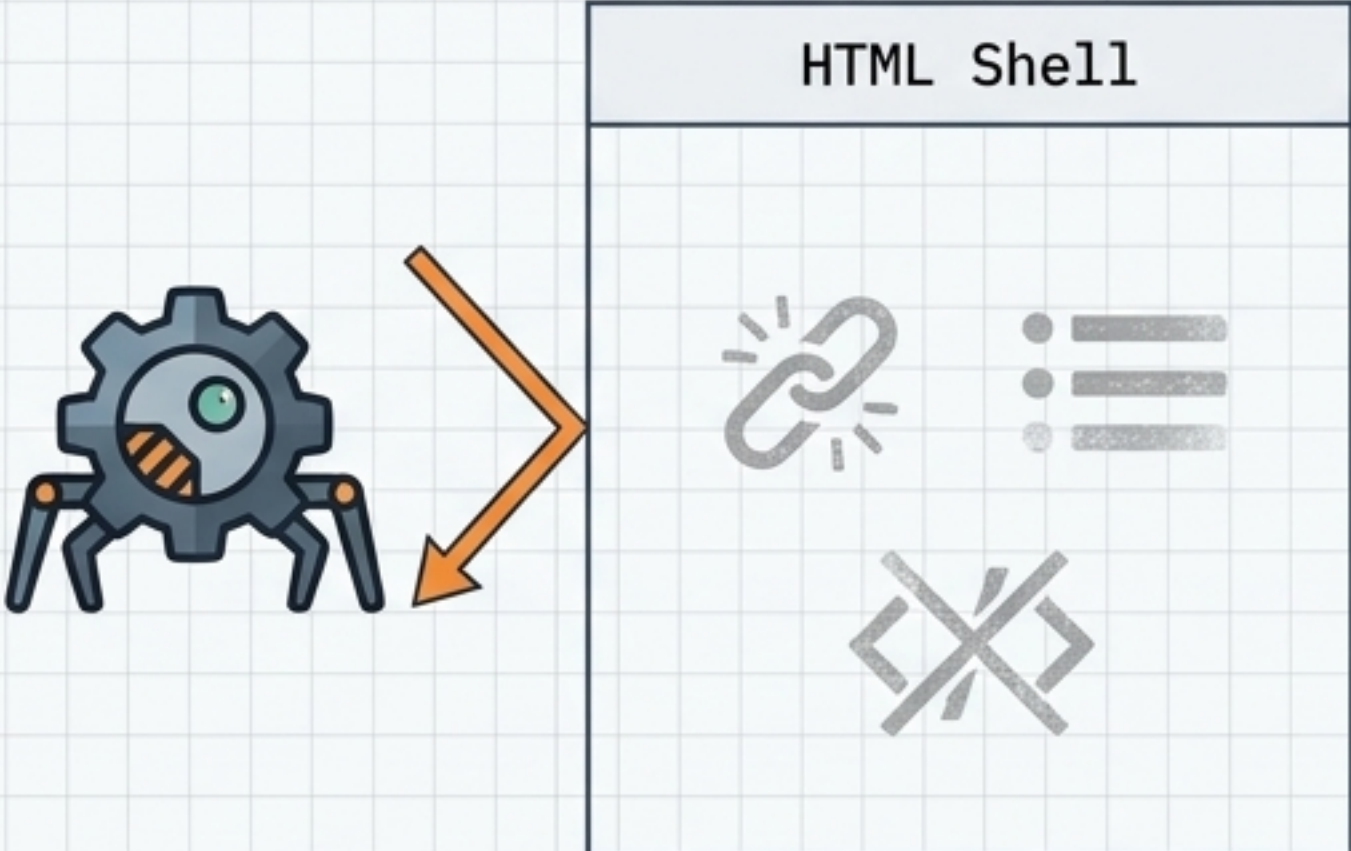
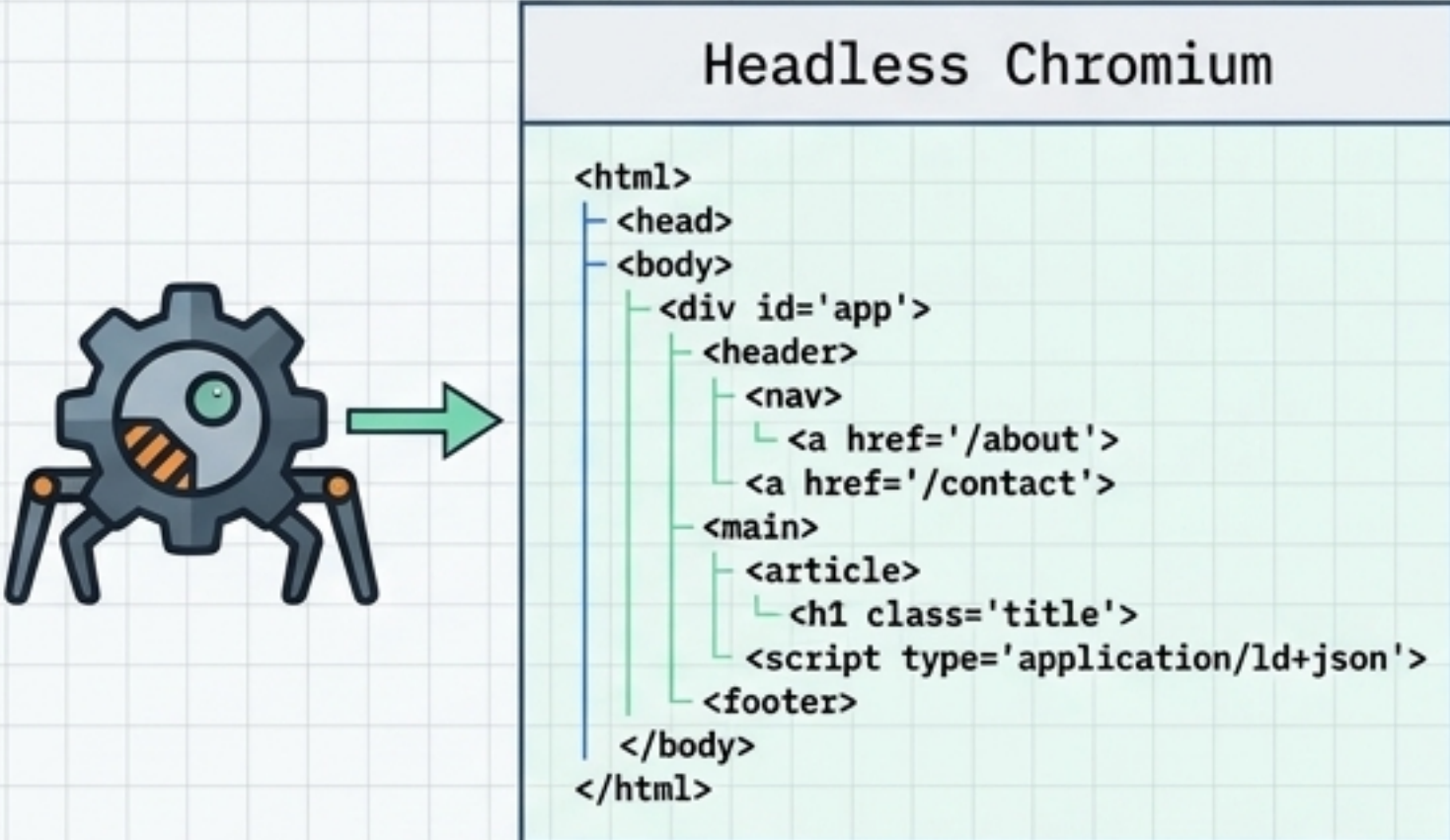
System Specs

Tool:	Screaming Frog SEO Spider (v24.1)
Free Tier Limit:	500 URLs
Pro License:	Unlimited capacity (\$279 USD/yr)

Crawl Mode Selection



The Render Mode Matrix: Don't crawl a blank shell.

Text Only (The Default)	JavaScript (The Requirement)
 <p data-bbox="806 609 1499 1388">HTML Shell</p>	 <pre data-bbox="2249 609 3105 1388"><html> <head> <body> <div id='app'> <header> <nav> </nav> <main> <article> <h1 class='title'> <script type='application/ld+json'> </article> </main> </div> </body> </html></pre>
<p data-bbox="239 1472 1482 1665">Mechanics: Parses raw HTML only. Silently misses JavaScript-injected canonicals, navigation links, and structured data.</p>	<p data-bbox="1772 1472 3105 1665">Mechanics: Executes client-side scripts. Mandatory for modern frontends to accurately capture the true page structure.</p>

The Dual User-Agent Setup: Simulating Googlebot.

HTTP Request User-Agent

Determines what headers are sent to the server.
Triggers dynamic rendering or mobile-specific content delivery.



Robots User-Agent

Determines which Disallow/Allow directives apply during the crawl.





Best Practice: Set both configurations to 'Googlebot' when auditing for organic visibility to ensure crawl-delay and block rules match reality.

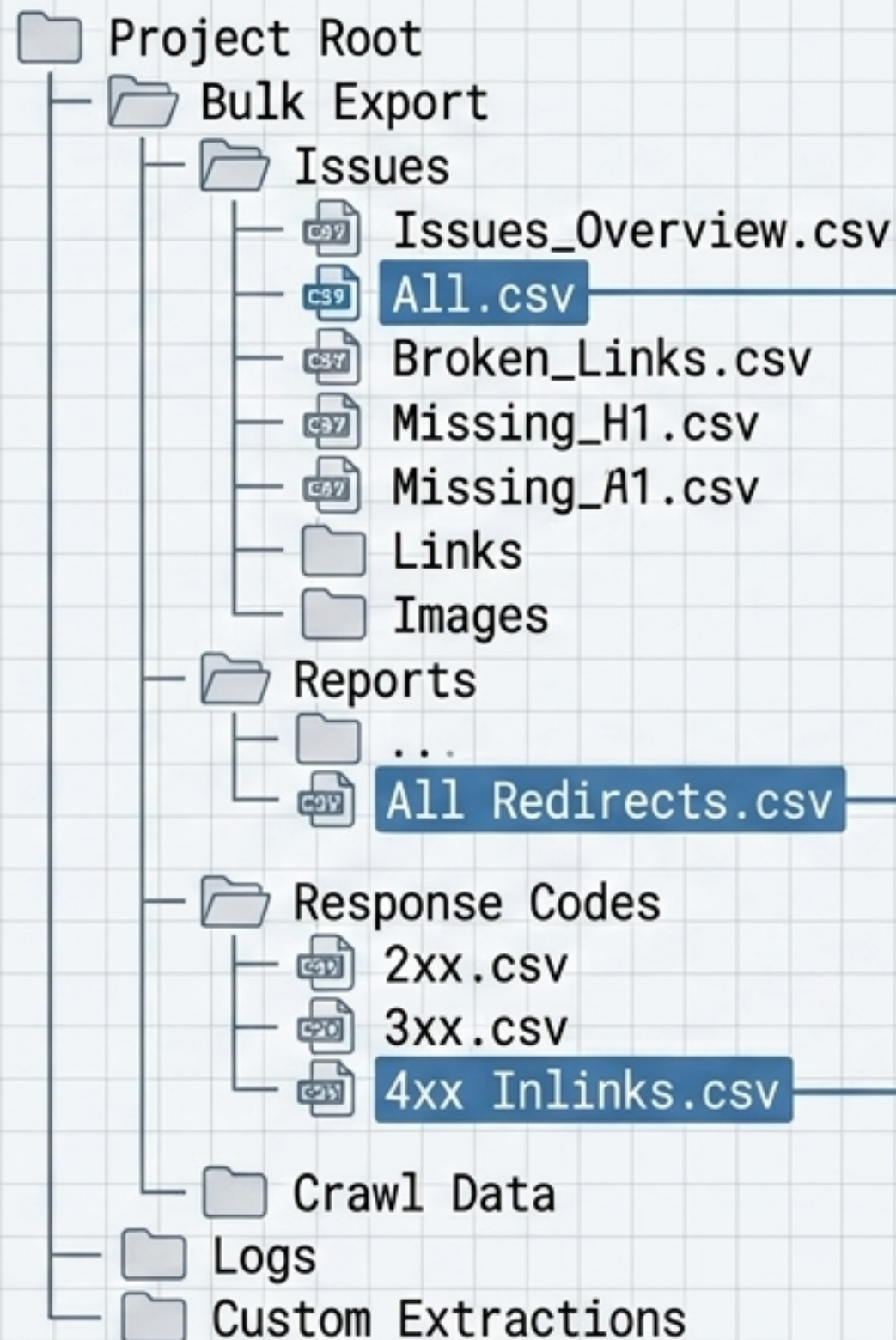
Custom Extraction: Scraping the hidden data layer.

```
Web Inspector – Source Code
<html>
  <head>
    <script>
      <metan src="/page".js</sp></script>
    </script>
    <link rel="canonical" href="...">
    <div id="data-container">
      var pageData = { "title": ... }
    </div>
    <div class="product-price" data-price="...">
    </div>
    <script>
      var pageData = {
        "title": "product",
        "JSON-LD schema blocks"
      }
    </script>
  </body>
</html>
```

LIMITS: Max 100 extractors per crawl / 1,000 total extractions.

	Method	Description
1	</> XPath	Best for structured HTML tags (e.g., canonical <link rel='canonical'> href values).
2	 CSS Path	Best for front-end visual elements and data-attributes.
3	Ax  Regex	Best for scraping inline script variables or JSON-LD schema blocks.

The Diagnostic Outputs: Three essential exports.



Path: "Bulk Export > Issues > All"

Purpose: "The master file. One CSV per issue category (Broken Links, Missing Metas, Low Content)."

Path: "Reports > All Redirects"

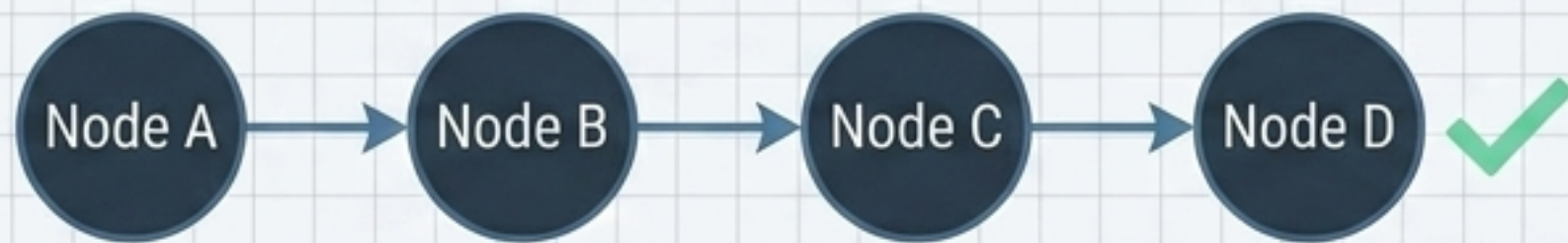
Purpose: "The path mapping. Shows hop counts, final destination codes, and critical loop identifiers."

Path: "Bulk Export > Response Codes > 4xx Inlinks"

Purpose: "The source locator. Identifies exactly which pages are linking to broken URLs so developers can actually fix them."

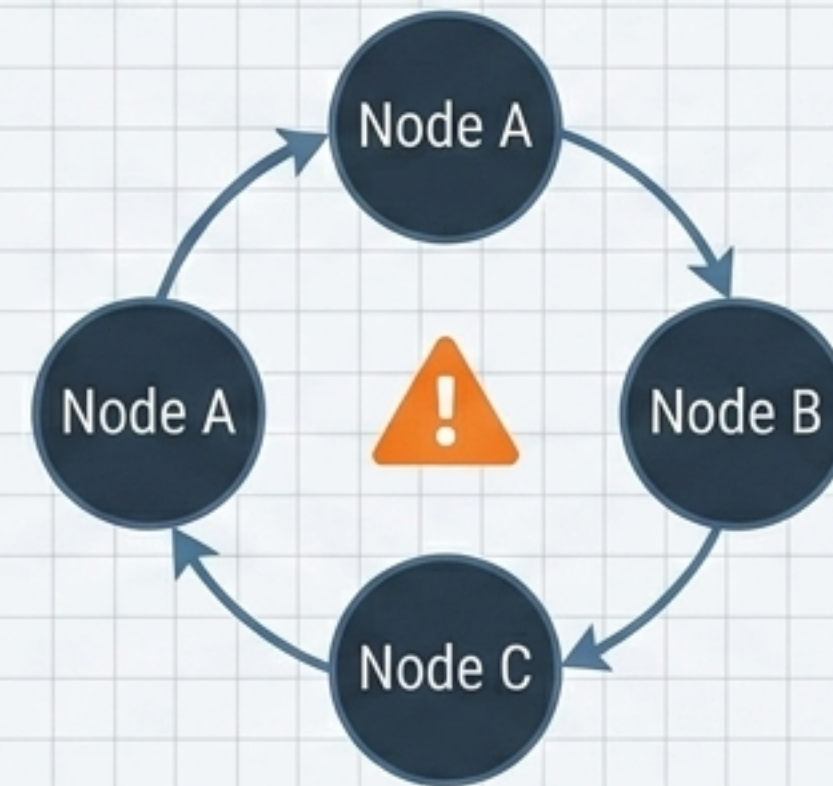
Redirects Diagnosed: The Chain vs. The Loop.

The Chain (Medium Severity / Warning)



Impact: Wastes crawl budget and dilutes link equity, but Googlebot eventually reaches the final 200 destination.

The Loop (High Severity / Error)

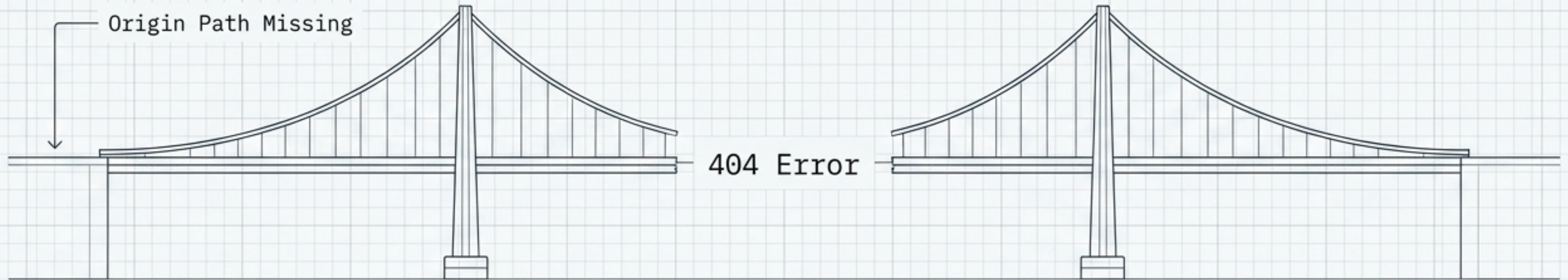


Impact: The 'Redirect Loop = True' flag. Googlebot is trapped and cannot index the page.

The Fix Sequence:

- 1) Resolve CMS circular reference.
- 2) Update inlinks.
- 3) Implement direct 301.

The 404 Paradox: You can't fix a destination without the origin.



The Pitfall:

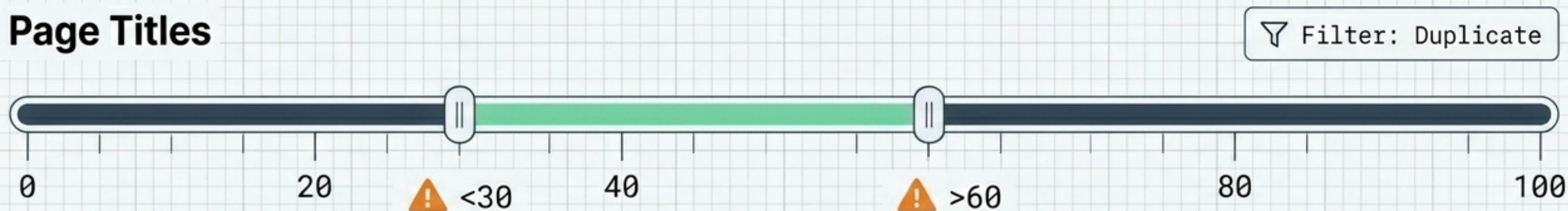
Exporting just the '4xx Client Error' list tells developers a page is dead, but it **leaves the broken link actively sitting on the site.**

The Solution:

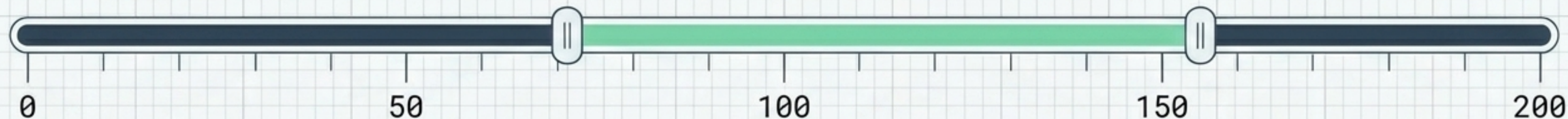
You must run the 'Response Codes > 4xx Inlinks' **export**. This provides the **exact source pages** that need their `<a href>` **tags updated** in the CMS.

Metadata Boundaries: Identifying duplicate and truncated signals.

Page Titles



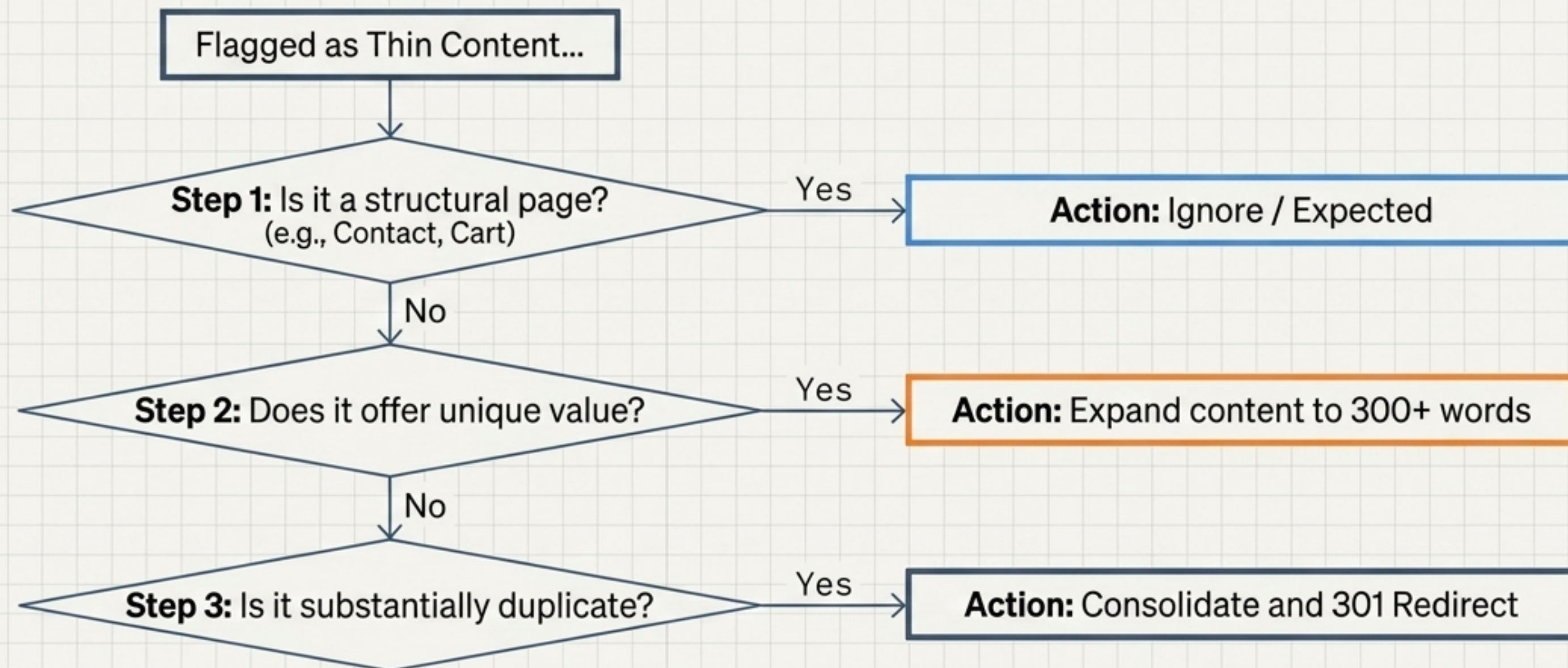
Meta Descriptions



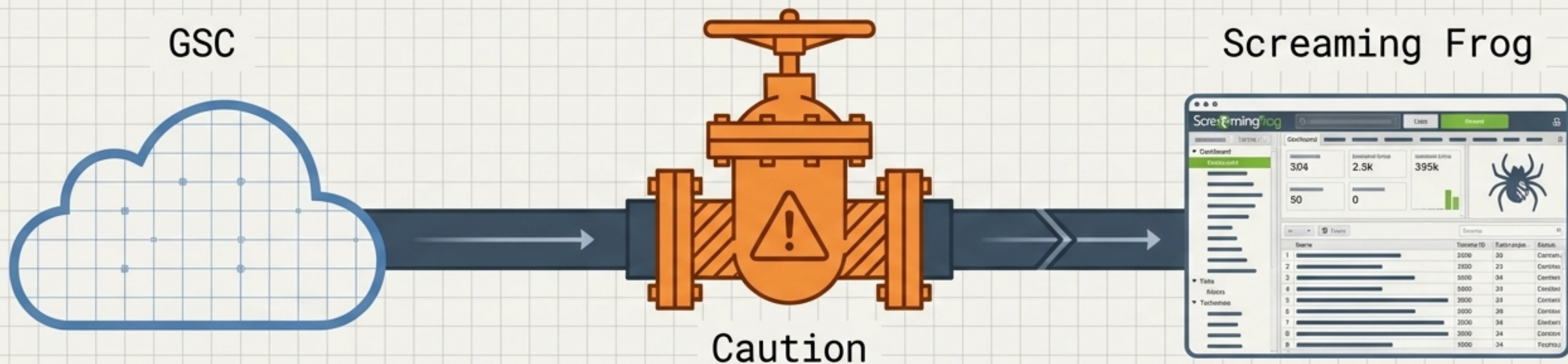
The 'Page Titles > Duplicate' filter surfaces **exact-match strings**. Duplicate titles across a site reduce ranking differentiation and confuse search engine clustering.

Thin Content: Context matters more than raw word count

Threshold Spectrum



Validation: Layering Google Search Console over crawl data.



The Integration:

Config > API Access > Google Search Console.
Pulls Indexing status, Coverage reasons, and Rich Result data.

The Bottleneck:

Google enforces a strict limit of 2,000 URLs per day per property.

The Workaround:

On an 85k+ URL site, do not run a full API crawl. Prioritize top pages by using **List Mode** with a curated URL set from GSC Search Analytics.

The GSC Indexing Translation Matrix

Status Code	Fix Priority	Root Meaning	Action Required
Crawled – currently not indexed	[PRIORITY: HIGH]	Content quality failure.	Improve content or consolidate duplicates.
Duplicate, Google chose different canonical	[PRIORITY: HIGH]	Disagreement on the primary URL.	Resolve conflicting canonical tags.
Discovered – currently not indexed	[PRIORITY: MED]	Crawl budget pressure.	Investigate server overload or site architecture limits.
Alternate page with proper canonical tag	[PRIORITY: NONE]	Expected behavior.	No action required.

A flat list is not a strategy: The triage mandate.

404 Not Found: /page1
Redirect Chain: /old-product
Missing Title Tag: /blog/post-123
Duplicate Content: /category/page2
Broken Link: /contact
Slow Load: /features
H1 Missing: /about
Image Alt Missing: /services
Broken Link: /contact
Broken At Missing: /about
Slow Views: /page
Image Link: /features
Missing Title Tag: /fzs
Broken Link: /wraps
404 Not Found: /page1
Duplicate Content: /category/page2
Duplicate Content: /category/page2
404 Loads Link: /contact
Missing Title Tag: /blog/post-123
Missing Missing: /fronse

$$\text{Traffic Impact} \times \text{Implementation Effort} = \text{Sprint Priority}$$

Scoring Traffic Impact (via GSC Analytics):

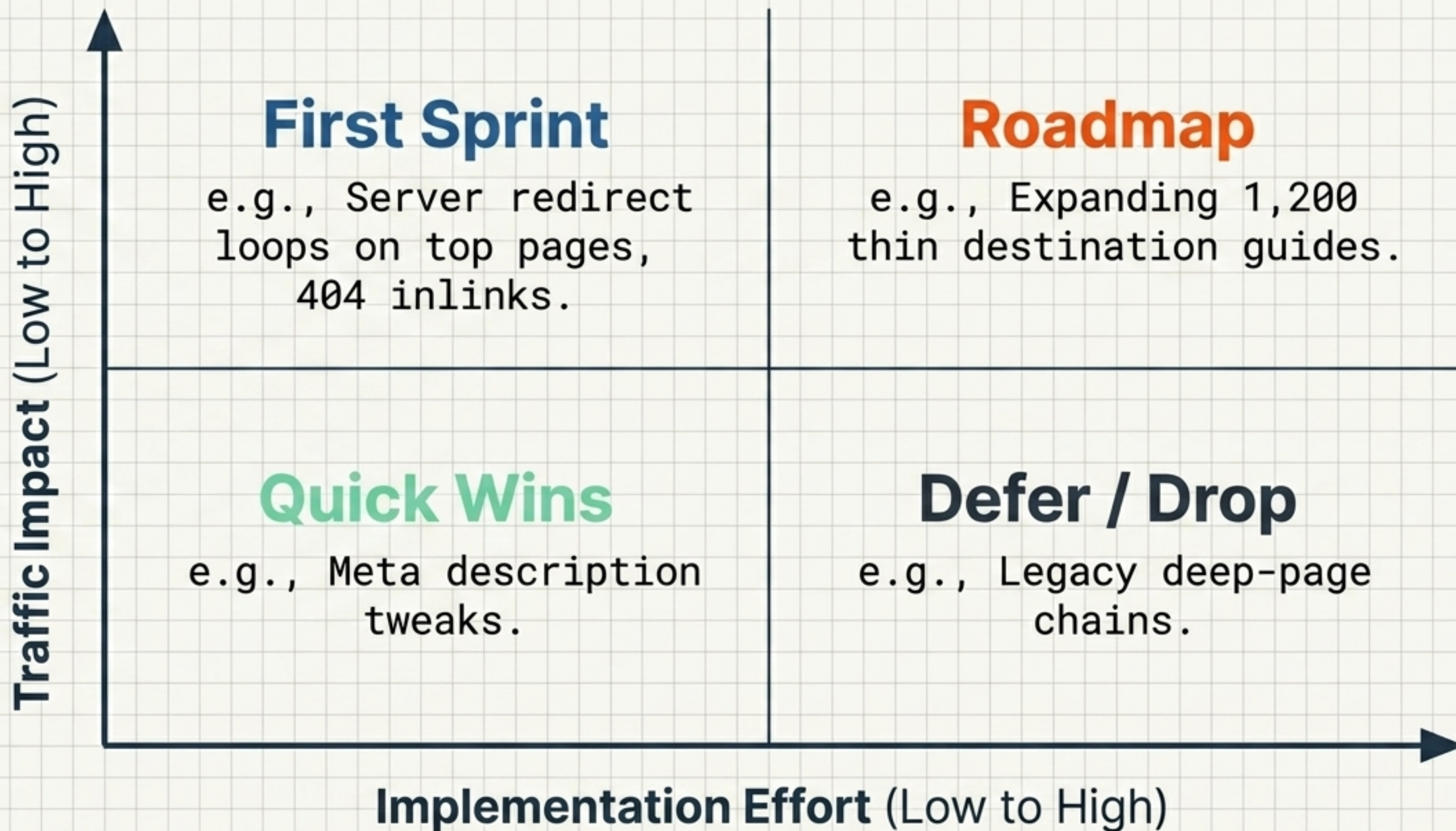
High: >500 impressions/month.

Medium: 50–500 impressions/month.

Low: <50 impressions/month.

Building a fix backlog in discovery order guarantees you will spend sprint time on **low-traffic title tags** while **high-traffic redirect loops** burn revenue.

The Prioritization Matrix: Plotting the fix backlog.



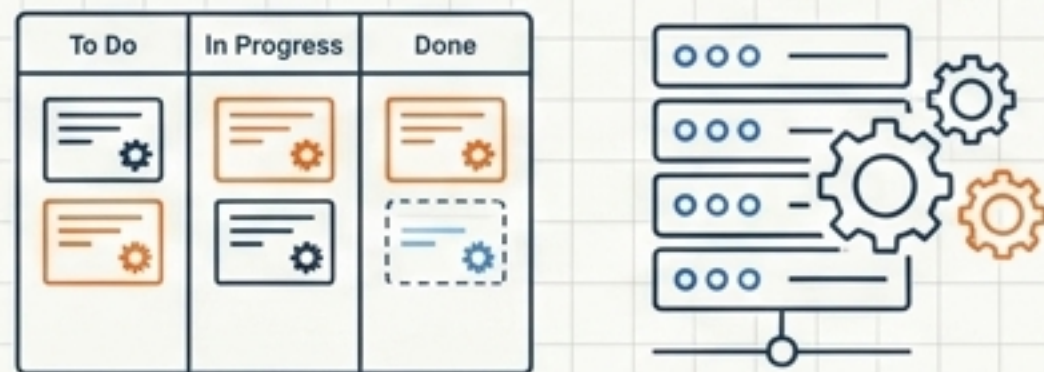
Assigning Ownership: Two lanes of execution.

Track 1: SEO-Team-Executable (No Dev Required)



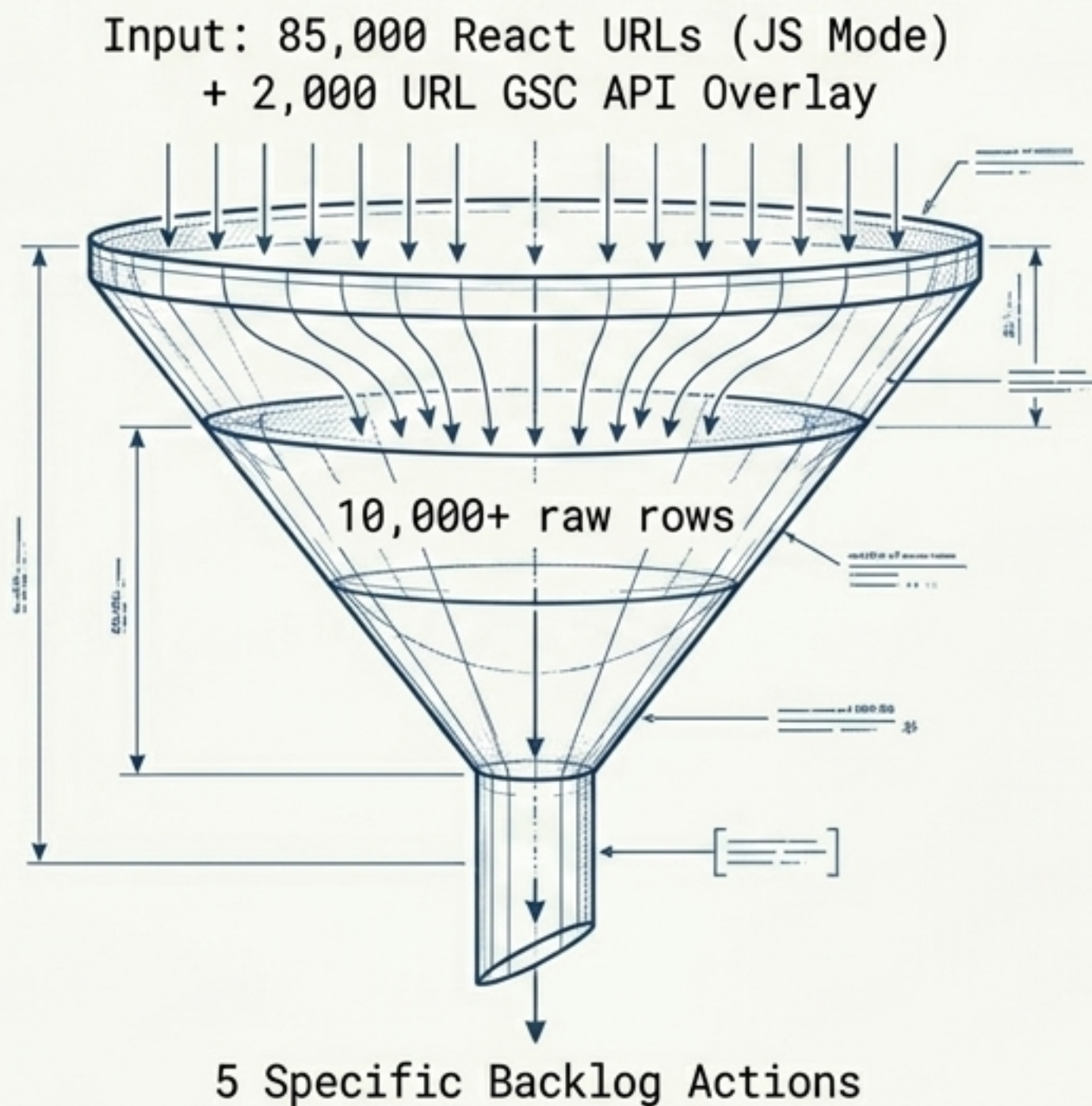
- Internal link corrections (fixing 4xx inlinks)
- Rewriting duplicate title tags
- Expanding thin content
- Updating meta descriptions

Track 2: Developer-Dependent (Sprint Scoped)



- Server-side 301 loop resolution
- Template-level H1 fixes
- React JavaScript render configuration
- Site-wide canonical logic

Case Study Synthesis: FlyAway Deals OTA Audit.



Top 3 Prioritized Fixes:

[High Impact / Low Effort]

Resolve circular redirect loop on /deals/summer (Dev).


[High Impact / Low Effort]


Update 47 broken inlinks pointing to top-50 hotel landing pages (Content/SEO).


[Med Impact / High Effort]


Expand 1,200 "Crawled - not indexed" destination guides currently under 140 words (Content/SEO).


The Auditor's Pre-Flight Checklist: Avoid these 5 traps.

 **[WARNING]** Don't use 'Text Only' render mode on React/Vue stacks.

 **[WARNING]** Don't misclassify a Redirect Loop (trapped) as a Redirect Chain (slow). Fix loops first.

 **[WARNING]** Don't export 404 errors without exporting the '4xx Inlinks' report.

 **[WARNING]** Don't apply the 200-word thin content flag blindly to structural pages (carts, confirmations).

 **[WARNING]** Don't hit the 2,000-URL GSC API limit blindly; prioritize via List Mode.