



Technical Case Study

Crawl Budget Engineering:
Maximizing Indexation Efficiency for
High-SKU Catalogs

Kristina Lichtenwald

SEO STRATEGIST & DATA ENGINEER
FOCUS: LOG FILE ANALYSIS, INDEXATION LOGIC, ROBOTS.TXT
GOVERNANCE, INTERNAL LINK DEPTH

The Challenge:

A rapidly expanding product catalog led to "Crawl Bloat." Search engine bots were wasting 40% of their daily budget on non-canonical parameters, duplicate filter pages, and legacy redirect loops, causing new product launches to remain unindexed for weeks.

The Objective:

Reclaim crawl budget by eliminating technical noise and engineering a "Priority Path" for search bots.

The "Zero-Waste" Approach

Step 1: Log File Forensic Audit:

Analyzed server logs to identify where Googlebot was spending time. Discovered high-hit rates on "Near-Duplicate" faceted navigation pages.

Step 2: Robots.txt & Directive Hardening:

Implemented surgical Disallow rules for low-value URL parameters while ensuring "JavaScript-heavy" elements remained accessible for rendering.

Step 3: Redirect Chain Remediation:

Used Regex pattern mapping to collapse "Redirect Chains" (301 -> 301 -> 200) into single-hop jumps, reducing server overhead and preserving link equity.

Crawl Metric	Legacy "Bloated" State	Engineered Efficiency State
<i>Indexation Latency</i>	7-14 Days for new SKUs	<24 Hours for new SKUs
<i>Crawl Waste</i>	40% (Parameters/Duplicates)	<5% (Verified Priority Pages)
<i>Crawl Depth</i>	7+ Clicks from home	<4 Clicks (Flattened IA)
<i>Outcome</i>	<i>Stale Search Results</i>	<i>Real-Time Catalog Sync</i>

Technical Execution

XML Sitemap Logic

Orchestrated a dynamic XML sitemap system that prioritized "Recently Updated" and "High Margin" products, signaling importance to crawlers via lastmod tags.

Internal Link Flattening

Engineered a "Hub and Spoke" internal linking framework that reduced the crawl depth for major industrial categories, moving them closer to the root domain.

Automation

Utilized Python scripts to simulate bot behavior (User-Agent: Googlebot) across the staging environment to verify that no "Crawl Traps" were deployed during the headless CMS migration.

Results & Business Impact

Improved Visibility

Within 60 days, indexation of the "New Arrivals" category increased by 85%.

Revenue Protection

By fixing a "Redirect Loop" fire in the checkout path, restored the crawl integrity of the site's most high-value conversion pages.

Stakeholder Alignment

Provided a "Crawl Health" dashboard in Looker Studio, allowing the engineering team to see the immediate impact of their code deployments on bot efficiency.

This and the other case studies represent Technical Proofs of Concept designed to demonstrate high-level SEO architecture, Python-driven automation, and advanced data modeling.

While the brand names and specific datasets are simulated to protect proprietary methodologies, each scenario is built on real-world logic, live SERP data, and enterprise-level growth frameworks. These case studies serve as a sandbox for testing zero-waste efficiency before deployment in live production environments.

None of the solutions of the case studies are "off the shelf." They are custom-built to eliminate waste and demonstrate some of my knowledge.