

# Cold Email Lead Pipeline

## Automated Sourcing, Verification & CRM Sync

**Stack:** n8n · Apify · Snov.io · Google Sheets · HubSpot

**Type:** Portfolio build, inspired by a real Upwork brief (Cold Email Operations Manager, 50K emails/month)

**Status:** Lead sourcing → verification → CRM sync fully built and tested. Sending (Instantly) and performance monitoring are a documented next phase, not yet built — see Scope below.

## The Challenge

High-volume B2B cold email programs live or die on one thing most people underestimate: the quality of what goes *into* the sender, not the sender itself. A tool like Instantly can send flawlessly and still tank a domain's reputation if it's fed unverified, duplicate, or poorly-targeted leads.

The brief this was built against called for exactly that upstream discipline: source leads from a target list, verify real contact emails before anything gets sent, avoid reprocessing the same lead twice, and land clean records in a CRM automatically — all without manual spreadsheet work.

## Architecture

Built using the same Receive → Normalize → Enrich → Decide → Route → Update framework used across this portfolio's other automation projects, so the logic stays consistent and auditable across builds.

| Stage                 | What happens                                                                                            | Tools                                                             |
|-----------------------|---------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------|
| Receive               | Scheduled trigger scrapes target company's employee list                                                | n8n Schedule Trigger + Apify (LinkedIn Company Employees Scraper) |
| Normalize             | Raw scrape output flattened into first name, last name, company                                         | n8n Edit Fields                                                   |
| Decide (dedup)        | Each lead checked against a tracking sheet; already-processed leads dropped before spending API credits | Google Sheets lookup + Merge + Filter                             |
| Enrich                | Company name resolved to a real domain, then used to find a verified work email                         | Snov.io (OAuth token → domain resolver → email finder)            |
| Decide (quality gate) | Leads with no findable email dropped rather than pushed downstream                                      | n8n Filter                                                        |
| Normalize (final)     | Clean, flat lead object built: email, name, company domain, LinkedIn URL                                | n8n Edit Fields                                                   |
| Route                 | Verified lead upserted as a CRM contact (safe to rerun, no duplicates)                                  | HubSpot                                                           |
| Update                | Lead marked processed so future runs skip it                                                            | Google Sheets append                                              |

~16 nodes end to end, fully tested against live API calls, not just wired up and assumed working.

## Tech Stack & Cost

| Tool          | Role                                   | Cost                                |
|---------------|----------------------------------------|-------------------------------------|
| n8n           | Orchestration                          | \$0 incremental (existing instance) |
| Apify         | Lead sourcing (LinkedIn scrape)        | Free tier (\$5/mo platform credit)  |
| Snov.io       | Domain resolution + email verification | Free tier (50 credits/month)        |
| Google Sheets | Dedup tracking log                     | Free                                |
| HubSpot       | CRM (contact storage)                  | Free tier                           |

**Total cost to build and test: \$0.** The original brief specified Make.com and AnyMailFinder; both were swapped for free-tier equivalents already in this stack (n8n replacing Make.com, Snov.io replacing AnyMailFinder after a sign-up blocker) without losing any functionality.

## Technical Challenges Solved

A few things came up mid-build that are worth flagging, since they're the kind of failure mode that isn't obvious until you're actually running real API calls:

- **Missing domain field.** The LinkedIn scraper returns rich profile data but no company website — only a LinkedIn slug. Solved by chaining in Snov's own name-to-domain resolver as an extra enrichment step.
- **Multi-domain companies break automatic resolution.** Testing against Google returned the wrong domain (goo.gle instead of google.com), since large companies often own dozens of domains. Solved by recognizing this as a test-data problem, not a pipeline bug, and re-testing against a single-domain company.
- **Silent async failures.** Two Snov.io endpoints process asynchronously (return a task handle, not an instant result) — naive sequential calls returned "in progress" instead of real data. Solved with a Wait node before polling for the result.
- **n8n's item-tracking breaks after external lookups.** Once a Google Sheets lookup node was inserted mid-pipeline, downstream nodes referencing earlier data by name threw "multiple matching items" errors, since n8n couldn't trace lineage through a node that can return a fabricated empty result. Solved with an explicit Merge node (combine by position) instead of relying on automatic item matching.
- **Graceful degradation over hard failures.** Leads with no resolvable domain would otherwise crash the whole batch. Configured with "continue on error" so one bad record routes to a separate branch instead of halting every other lead in the run.

## Results (Test Run)

- 9 employee leads scraped from a single target company in one run
- 4 of 9 (44%) resolved to a verified, deliverable-looking email address — a realistic hit rate for this kind of lookup, not a cherry-picked number
- All 4 verified leads landed correctly in HubSpot as upserted contacts, confirmed via HubSpot's own returned contact IDs
- Zero duplicate processing on rerun, confirmed via the dedup check

## Scope Note

This build covers lead sourcing, verification, deduplication, and CRM handoff — the part of a cold email operation most likely to silently damage deliverability if done carelessly. It does not yet include the sending layer (Instantly) or performance monitoring/optimization, which were part of the original brief but require a live sending account to build and test meaningfully. Documented as the natural next phase.