

Building a multi-source hiring intelligence engine with canonical normalization and change tracking

A Python acquisition layer that converts heterogeneous applicant-tracking systems into dependable, comparable job-market signals.

STATUS	TEAM	DEMO	SOURCE
Beta · Source reviewed · Selected tests passing	Solo engineering project	Fixture-based technical demo planned	Private repository

```
bash - pytest tests/

anass@workstation:~/Documents/Scrapers/ats_scrapers_v2$ pytest tests/test_normalizer.py
tests/test_delta.py tests/test_hvi.py -v

===== test session starts =====
platform linux -- Python 3.14.5, pytest-8.4.2, pluggy-1.6.0
rootdir: /home/anass/Documents/Scrapers/ats_scrapers_v2
configfile: pyproject.toml
plugins: cov-7.1.0, respx-0.23.1, mock-3.15.1, xdist-3.8.0, asyncio-0.26.0
collected 68 items

tests/test_normalizer.py::TestNormalizeTitle::test_basic_title PASSED [ 10%]
tests/test_normalizer.py::TestNormalizeTitle::test_senior_detection PASSED [ 20%]
tests/test_normalizer.py::TestNormalizeLocation::test_remote_detection PASSED [ 35%]
tests/test_normalizer.py::TestParseSalary::test_range_k PASSED [ 50%]
tests/test_delta.py::TestDeltaProcessor::test_empty_incoming PASSED [ 65%]
tests/test_delta.py::TestDeltaProcessor::test_all_new_jobs PASSED [ 75%]
tests/test_delta.py::TestDeltaProcessor::test_conditional_fetch_skip PASSED [ 85%]
tests/test_hvi.py::TestHVISignificance::test_mild_acceleration PASSED [ 92%]
tests/test_hvi.py::TestHVISignificance::test_strong_acceleration PASSED [100%]

===== warnings summary =====
tests/test_normalizer.py::TestNormalizeTitle::test_basic_title: DeprecationWarning

68 passed, 19 warnings in 1.24s
```

Primary visual evidence

ANASS BENAMEUR · FULL-STACK PRODUCT ENGINEER

The product, the problem and the proof

Every applicant-tracking system represents the same idea — a job posting — differently: different APIs, different HTML, different pagination, different field names for the same concept. This engine puts ten of them behind a shared adapter interface, converts everything into one canonical job model, and computes the delta between snapshots rather than re-ingesting the same postings as new data every run.

10 ADAPTER IMPLEMENTATIONS	24 TEST MODULES	68 TESTS PASSED IN CAPTURED RUN	1.24 s CAPTURED SELECTED-SUITE RUNTIME
--------------------------------------	---------------------------	--	---

Context

Structured as a proper Python package — Python 3.11+, database schemas, orchestration scripts, observability hooks — with dedicated modules for rate limiting, proxy health, normalization, revision history and hiring-velocity scoring, aimed at recurring monitoring rather than a one-off export. It is the acquisition core that also powers the job-market side of the larger Elevya Signal Intelligence platform elsewhere in this portfolio, built first here as a standalone, reusable package.

My role and ownership

- Adapter and registry architecture
- Normalization and delta processing
- Rate limiting, retries and proxy-health controls
- PostgreSQL schemas and revision history
- Automated tests for core logic and sources

Personal contribution

Adapters, registry, canonical model, state comparison, revisions, network controls and tests.

Why the existing workflow needed a system

A title with a requisition code baked in, a location string that mixes city and remote-work signals in no consistent order, a salary range written as free text, a contract type spelled four different ways across four platforms — none of that is usable for comparison until it's normalized. And without snapshot-to-snapshot comparison, "new data" is really just the same postings re-ingested on every run.

Before and after

BEFORE

- One-off source-specific scripts
- Different field names and types per ATS
- Repeated processing when boards have not changed
- No durable record of added, changed or removed jobs
- No company-level hiring momentum signal

AFTER

- Adapter registry with source-specific clients
- Canonical title, location, contract, salary and function fields
- Conditional fetch and rate-limit controls where supported
- Delta and revision layers for lifecycle tracking
- High-Velocity Indicator for acceleration classification

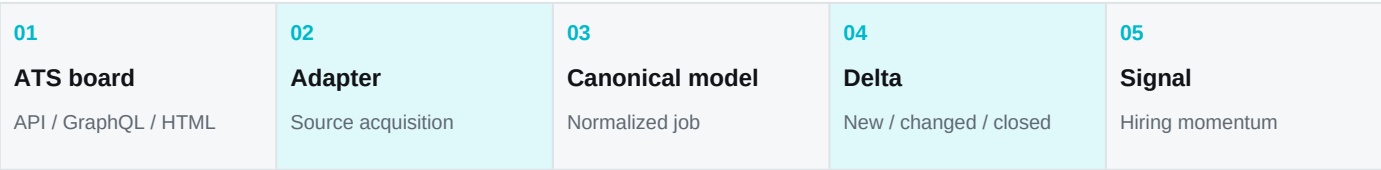
Meaningful result

Ten ATS integrations converge on one canonical job and change model.

How the system works

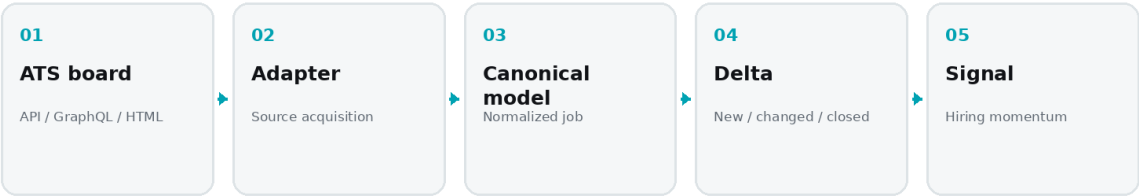
Ten source adapters — Greenhouse, Lever, Ashby, Workable, Personio, BambooHR, Recruitee, SmartRecruiters, Teamtailor and France Travail — sit behind one registry and shared orchestration, so adding an eleventh platform means writing one adapter, not touching the pipeline that runs all ten. Every record, regardless of source, lands in the same canonical schema.

Operational workflow



SYSTEM ARCHITECTURE

ATS Intelligence Engine



Evidence-led architecture diagram · generated from the verified project workflow

A delta module compares each new snapshot against the last known state per source, so only genuine changes generate a revision record instead of re-writing everything every run. A hiring-velocity module turns the resulting change history into a company-level acceleration signal — the piece that makes ten normalized feeds worth more than ten spreadsheets, because it answers "who's hiring faster than last month," not just "who's hiring." The pipeline in sequence:

Architecture and decisions

The package targets Python 3.11+ and uses asyncio, httpx, psycopg and Selectolax. The adapter interface is strict enough that the orchestrator never needs to know which platform it's talking to, at the cost of upfront design time — a cost that pays back every time a new ATS gets added, since the tenth adapter took a fraction of the effort the first one did.

Critical engineering decisions

DECISION 01 Thin adapters, shared policy Platform-specific parsing stays isolated while retries, rate limits, metrics and database behavior improve once for the whole engine.	DECISION 02 Normalize before downstream use A canonical job record converts inconsistent titles, locations, contracts and salaries into fields that support analytics.
DECISION 03 Track revisions, not only current rows Delta processing distinguishes new, changed and closed jobs and preserves evidence for hiring momentum.	DECISION 04 Test contracts and edge cases The repository includes tests for infrastructure and adapters; the supplied capture shows 68 passing tests with 19 warnings.

Technology stack

Python 3.11+	asyncio	httpx	PostgreSQL	Adapters	Delta / revisions
--------------	---------	-------	------------	----------	-------------------

Evidence 1 of 2

```
bash - pytest tests/

anass@workstation:~/Documents/Scrapers/ats_scrapers_v2$ pytest tests/test_normalizer.py
tests/test_delta.py tests/test_hvi.py -v

===== test session starts =====
platform linux -- Python 3.14.5, pytest-8.4.2, pluggy-1.6.0
rootdir: /home/anass/Documents/Scrapers/ats_scrapers_v2
configfile: pyproject.toml
plugins: cov-7.1.0, respx-0.23.1, mock-3.15.1, xdist-3.8.0, asyncio-0.26.0
collected 68 items

tests/test_normalizer.py::TestNormalizeTitle::test_basic_title PASSED [ 10%]
tests/test_normalizer.py::TestNormalizeTitle::test_senior_detection PASSED [ 20%]
tests/test_normalizer.py::TestNormalizeLocation::test_remote_detection PASSED [ 35%]
tests/test_normalizer.py::TestParseSalary::test_range_k PASSED [ 50%]
tests/test_delta.py::TestDeltaProcessor::test_empty_incoming PASSED [ 65%]
tests/test_delta.py::TestDeltaProcessor::test_all_new_jobs PASSED [ 75%]
tests/test_delta.py::TestDeltaProcessor::test_conditional_fetch_skip PASSED [ 85%]
tests/test_hvi.py::TestHVISignificance::test_mild_acceleration PASSED [ 92%]
tests/test_hvi.py::TestHVISignificance::test_strong_acceleration PASSED [100%]

===== warnings summary =====
tests/test_normalizer.py::TestNormalizeTitle::test_basic_title: DeprecationWarning

68 passed, 19 warnings in 1.24s
```

Selected automated suite: 68 tests passed with 19 warnings in 1.24 seconds.

Evidence 2 of 2

ADAPTER COVERAGE · SOURCE-REVIEWED

Ten source adapters behind one canonical job model

01 Ashby adapter → canonical record	02 BambooHR adapter → canonical record	03 France Travail adapter → canonical record	04 Greenhouse adapter → canonical record	05 Lever adapter → canonical record
06 Personio adapter → canonical record	07 Recruitee adapter → canonical record	08 SmartRecruiters adapter → canonical record	09 Teamtailor adapter → canonical record	10 Workable adapter → canonical record

The public case study claims only adapters identified in the supplied repository.

Source-reviewed adapter coverage behind the canonical model.

Value, proof and honest limitations

Operational impact

Ten adapter source files, 24 test modules covering infrastructure and platform clients, and PostgreSQL schemas for core records, revisions and cache with optimized indexes — this is the acquisition layer that made the larger Signal Intelligence platform possible to build in the time it took, because the hardest normalization problem was already solved and tested here first.

Verified evidence

- 10 adapter source files identified
- 24 test modules covering infrastructure and platform clients
- Captured selected run: 68 passed, 19 warnings in 1.24 seconds
- PostgreSQL schemas for core records, revisions, cache and optimized indexes
- CLI scripts for sweeps, exports, HVI backfill and classifier training

Current limitations

- No live multi-platform sweep was captured for this portfolio package.
- The screenshot proves selected tests, not every production integration or credential path.
- An optimization percentage in source documentation lacks a benchmark artifact and is not repeated here.
- Warnings remain in the captured suite and should be removed or explicitly accepted.
- A public demo should use fixtures or cached responses.

Current status

STATUS	TEAM	DEMO	SOURCE
Beta · Source reviewed · Selected tests passing	Solo engineering project	Fixture-based technical demo planned	Private repository

WHY THIS PROJECT MATTERS

Ten ATS integrations converge on one canonical job and change model.

This case study is designed to be inspectable: strong claims are connected to screenshots, source evidence or explicitly stated limitations.