

# Alion Data Catalog Overview

Technology hiring data collected directly from employer applicant tracking systems and career pages. Coverage snapshot as of 2026-09-21. Company-level business data only; no personal data of candidates.

## 1. What Alion is

Alion operates a continuous crawler over employer job boards and career pages worldwide. Every posting is collected from the employer's own publication (ATS board, ATS API or schema.org career page), re-verified against its source until it disappears, and normalized into a single taxonomy of roles, technologies, seniority levels, locations and compensation. Company records are compiled from company websites, public filings and the company's own hiring activity. The result is a job-level and company-level view of technology hiring that is fresh (daily), provenance-tracked (source URL and timestamps on every record) and enriched with derived signals that raw posting scrapes do not carry: posting lifecycle and repost detection, stale/ghost scoring, salary estimates, remote scope and hiring geography, technology adoption over time.

## 2. Datasets

| Dataset                                                                                                                                                                | Unit            | Volume (2026-09-21)                                                      | Refresh                                                                        | History                                                                                                                                        | Sample                                                                                                      |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------|--------------------------------------------------------------------------|--------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------|
| <b>Global Tech Job Postings</b><br>Normalized postings with full text, taxonomy, compensation, geography, lifecycle and trust fields                                   | posting         | 655,259 live<br>274,911 new / 30d<br>244,513 closed / 30d                | daily (ATS boards re-verified daily, high-velocity boards several times a day) | closed postings retained with close dates; posting-level open/close/reopen events since July 2026; first-seen dates in the corpus back to 2020 | CSV, 2,500 rows                                                                                             |
| <b>Company Hiring Signals &amp; Firmographics</b><br>Employer-level hiring velocity, technology adoption, geography, pay level, trust, funding/revenue where available | company         | 943,548 companies<br>32,639 hiring now                                   | daily; weekly hiring snapshots                                                 | weekly snapshots since 2026-07-20; technology first-seen/last-seen dates; 2,907,550 dated growth events                                        | CSV, 1,500 rows                                                                                             |
| <b>Posting Lifecycle &amp; Trust Signals</b><br>Open/close events, days open, repost detection, stale and ghost scoring, apply-channel type                            | posting / event | 385,191 close events<br>109,925 reopen events<br>327,183 scored postings | daily                                                                          | events since 2026-07-23                                                                                                                        | included in the postings sample (status, closed at, days open, repost_count, trust_level, lifecycle_events) |

|                                                                                                           |                |                                                                                                          |         |                                                                                   |                  |
|-----------------------------------------------------------------------------------------------------------|----------------|----------------------------------------------------------------------------------------------------------|---------|-----------------------------------------------------------------------------------|------------------|
| <b>Tech Salary Benchmarks</b><br>Percentiles of stated and estimated pay by role, seniority and geography | benchmark cell | 213,542 live postings with stated pay<br>343,293 with model estimate                                     | daily   | rolling window (live + 180 days of closed postings); monthly snapshots on request | CSV, 4,094 cells |
| <b>Taxonomies</b><br>Delivered with every dataset                                                         | reference      | 3,549 technologies<br>1,198 roles in 21 families<br>709 industry markets<br>22,579 cities, 236 countries | monthly | versioned                                                                         | on request       |

### 3. Coverage snapshot: postings

|                                               |                                              |                                            |
|-----------------------------------------------|----------------------------------------------|--------------------------------------------|
| <b>655,259</b><br>live postings               | <b>32,639</b><br>employers hiring now        | <b>78,302</b><br>new in the last 7 days    |
| <b>274,911</b><br>new in the last 30 days     | <b>244,513</b><br>closed in the last 30 days | <b>199</b><br>countries with live postings |
| <b>1,310</b><br>cities with 50+ live postings |                                              |                                            |

#### Field coverage on live postings

| Field                                                | Postings | Share |
|------------------------------------------------------|----------|-------|
| Role mapped to taxonomy                              | 655,136  | 100%  |
| Seniority resolved                                   | 428,491  | 65.4% |
| Technologies tagged (1,811,749 tags, 2,407 distinct) | 352,957  | 53.9% |
| Stated salary, normalized to annual USD              | 213,542  | 32.6% |
| Model salary estimate (postings without stated pay)  | 343,293  | 52.4% |
| Remote                                               | 81,230   | 12.4% |
| Hybrid                                               | 112,636  | 17.2% |
| On-site                                              | 461,370  | 70.4% |
| Visa sponsorship stated                              | 3,567    | 0.5%  |
| Equity mentioned                                     | 46,828   | 7.1%  |

#### Top countries (office location of live postings)

| Country        | Live postings |
|----------------|---------------|
| United States  | 279,240       |
| Germany        | 37,535        |
| India          | 31,401        |
| United Kingdom | 30,758        |
| Canada         | 18,965        |
| France         | 15,120        |
| Singapore      | 10,430        |
| Netherlands    | 9,203         |
| Australia      | 9,138         |
| Spain          | 8,157         |
| Poland         | 8,046         |
| Mexico         | 6,272         |
| China          | 5,788         |
| Malaysia       | 5,462         |
| Philippines    | 5,201         |

About 20% of postings are remote or carry no office location and are not counted by country. Strongest coverage: United States, India,

|                                  |         |       |
|----------------------------------|---------|-------|
| Trust scored<br>(stale/ghost/ok) | 248,566 | 37.9% |
|----------------------------------|---------|-------|

United Kingdom, Canada, EU, Singapore.

#### Source type

| Source                                       | Live postings | Share |
|----------------------------------------------|---------------|-------|
| Employer ATS boards and APIs                 | 643,618       | 98.2% |
| Employer career pages<br>(schema.org / HTML) | 5,952         | 0.9%  |
| Partner feeds and APIs                       | 5,689         | 0.9%  |

#### Role families

| Family                 | Live postings |
|------------------------|---------------|
| Field & Service        | 81,257        |
| Sales                  | 75,627        |
| Industrial Engineering | 56,664        |
| Backend                | 42,079        |
| Management             | 33,839        |
| Operations             | 33,279        |
| Finance                | 32,609        |
| Marketing              | 29,091        |
| HR                     | 28,341        |
| Early Careers          | 27,141        |
| Administrative         | 26,152        |
| Customer Success       | 17,258        |

#### Seniority

| Level            | Live postings |
|------------------|---------------|
| Senior           | 146,803       |
| Staff +          | 29,332        |
| Lead             | 51,014        |
| Head / Executive | 30,396        |
| Middle           | 64,059        |
| Junior           | 85,262        |
| Intern           | 21,619        |
| Not resolved     | 226,775       |

#### Employment type

| Type       | Live postings |
|------------|---------------|
| Full-time  | 413,685       |
| Contractor | 27,052        |
| Internship | 22,105        |
| Part-time  | 33,732        |

#### ATS platforms (employers with live postings)

| ATS             | Employers | Live postings |
|-----------------|-----------|---------------|
| Workday         | 902       | 181,558       |
| Greenhouse      | 4,509     | 99,154        |
| SmartRecruiters | 1,269     | 77,514        |
| Workable        | 3,169     | 60,137        |
| Ashby           | 3,261     | 47,492        |
| Lever           | 1,651     | 46,752        |
| Oracle          | 173       | 17,623        |
| Recruitree      | 1,493     | 12,525        |
| Teamtai         | 1,313     | 11,994        |
| ATS #36         | 7,018     | 11,558        |

|                    |        |            |         |              |       |        |
|--------------------|--------|------------|---------|--------------|-------|--------|
| Product            | 15,254 | Freelance  | 2,163   | Brezy        | 998   | 11,212 |
| DevOps             | 15,184 | Co-founder | 98      | Eightfold    | 24    | 10,829 |
| Legal              | 13,187 | Not stated | 156,425 | Schem a.org  | 279   | 9,900  |
| AI/ML              | 13,157 |            |         | career pages |       |        |
| Solutions          | 12,188 |            |         | Personio     | 1,202 | 7,934  |
| Data Science       | 10,959 |            |         | ATS #11      | 394   | 7,661  |
| Science & Research | 10,089 |            |         | Rippling     | 1,054 | 6,761  |
| Analytics          | 10,058 |            |         |              |       |        |

60+ ATS platforms supported in total.

#### 4. Coverage snapshot: companies

| Attribute                                                                      | Companies | Note                                                      |
|--------------------------------------------------------------------------------|-----------|-----------------------------------------------------------|
| Company records                                                                | 943,548   | technology employers and companies with technology hiring |
| Hiring now (live postings)                                                     | 32,639    | hiring signals are computed for these                     |
| Industry classified                                                            | 152,402   | 651-market taxonomy, two levels                           |
| Description                                                                    | 338,964   | from company website                                      |
| Office locations                                                               | 123,166   | HQ flag, city and country                                 |
| Founded year                                                                   | 87,495    |                                                           |
| ATS platform identified                                                        | 42,948    |                                                           |
| Technology stack from hiring (855,939 company-technology rows)                 | 36,280    | first-seen / last-seen dates per technology               |
| Weekly hiring snapshots since 2026-07-20                                       | 58,974    | 188,770 weekly rows                                       |
| Hiring signal scorecards (demand, pay, trust, spikes)                          | 132,947   |                                                           |
| Dated growth events (geo expansion, tech added/removed, hiring spike, funding) | 2,907,550 | events, not companies                                     |
| Social profiles (LinkedIn, GitHub, X, company-level)                           | 203,236   |                                                           |

|                                     |        |                                                                                              |
|-------------------------------------|--------|----------------------------------------------------------------------------------------------|
| Size band (evidence-based estimate) | 49,116 | published only above a confidence threshold; higher coverage among actively hiring companies |
| Funding rounds                      | 111    | sparse, venture-backed subset                                                                |
| Revenue from public filings         | 164    | sparse, listed companies                                                                     |
| Flagged staffing agencies           | 3,831  | separable from direct employers                                                              |

## 5. Freshness, history and verification

- **Collection cadence.** Active employer boards are re-scanned daily; high-velocity boards several times a day. New employers are discovered continuously from company domain lists and ATS detection.
- **Verification.** A posting stays *live* only while it is present on its source board. When it disappears, a *close* event is written with the timestamp; re-appearance writes a *reopen* event. Postings that stop being verifiable are *expired*. Close events have been journaled since 2026-07-23 (385,191 events so far).
- **Repost and trust scoring.** The same vacancy re-published under a new identifier is linked to its predecessor; postings open far beyond typical time-to-fill for their role, or belonging to employers with a company-wide stale pattern, are scored *stale* or *ghost*. Today 327,183 postings carry a score (128,010 ok, 112,668 stale, 86,505 ghost); coverage grows as the corpus accumulates observation history.
- **History.** Closed postings are retained (255,967 on file) with close dates. Weekly company hiring snapshots since 2026-07-20. Technology adoption history through per-company first-seen and last-seen dates.
- **Provenance.** Every record keeps its source URL, first-seen and last-verified timestamps.

## 6. Known limitations

- Focus is technology hiring: engineering, data, AI/ML, product, design, security, IT and adjacent roles. Non-technical postings of the same employers are only partially covered.
- Stated salaries are parsed from free text in many languages and currencies; occasional parsing errors survive validation. Consumers of compensation data should apply the `stated_n` thresholds in the benchmarks and treat single-posting figures accordingly.
- Company size bands are model estimates published only above a confidence threshold; funding and revenue fields cover a small subset of companies.
- Trust scoring covers about a third of live postings today; younger postings are unscored by design.
- Posting-level lifecycle events start in July 2026; earlier history is limited to first-seen dates and retained closed postings.
- Data from regional job boards (Russia/CIS) is excluded from standard delivery and available separately on request.

## 7. Delivery, licensing, compliance

### Delivery

- REST API (JSON), filtered by country, technology, role, seniority, company or first-seen date.
- Bulk files: daily full refresh or daily deltas, CSV / JSON Lines, delivered by HTTPS download, S3 or SFTP push.
- Custom extracts and company lists on request; taxonomy files included.
- Free sample files (this catalog) and a trial period with full fields on a limited slice.

### Licensing

- Monthly or annual subscription; tiers by dataset, geography and refresh frequency.
- Internal-use licence by default; redistribution and derived-product licences on request.
- Data licence agreement and, where relevant, a data processing agreement available on request.

### Compliance

- Company-level business data collected from publications the employer made public on its own board or website.
- No candidate profiles, resumes or any personal data of job seekers are collected or distributed.
- No social networks or professional networking sites are scraped.
- Rate-limited collection; source URL and timestamps retained per record for audit.

## 8. Sample files

| File                                                             | Rows  | Content                                                                                                                                                                                                                                         |
|------------------------------------------------------------------|-------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <a href="#">alion_sample_job_postings_20260921.csv</a>           | 2,500 | 2,200 live postings (max 3 per employer, all ages) and 300 postings closed in the last 30 days; 55 fields including lifecycle, trust, stated and estimated pay, technologies, geography, apply channel and a 400-character description excerpt. |
| <a href="#">alion_sample_company_hiring_signals_20260921.csv</a> | 1,500 | 1,000 largest technology hirers plus 500 random employers with 3+ live postings; 63 fields of firmographics, hiring velocity, weekly series, technology adoption, pay, trust and growth events.                                                 |
| <a href="#">alion_sample_salary_benchmarks_20260921.csv</a>      | 4,094 | Benchmark cells (role and role family x seniority x country / Remote / All) computed from 542,342 observations; stated percentiles at $n \geq 8$ , estimate percentiles at $n \geq 30$ .                                                        |

All figures in this document are computed from the production corpus on 2026-09-21 and exclude regional job board data. Alion, [alion.io](#).