

Alion Data Catalog Overview

Technology hiring data collected directly from employer applicant tracking systems and career pages. Coverage snapshot as of 2026-09-04. 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 (2026-09-04)	Refresh	History	Sample
Global Tech Job Postings Normalized postings with full text, taxonomy, compensation, geography, lifecycle and trust fields	posting	121,355 live 48,653 new / 30d 34,620 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
Company Hiring Signals & Firmographics Employer-level hiring velocity, technology adoption, geography, pay level, trust, funding/revenue where available	company	93,957 companies 7,855 hiring now	daily; weekly hiring snapshots	weekly snapshots since 2026-07-20; technology first-seen/last-seen dates; 287,085 dated growth events	CSV, 1,500 rows
Posting Lifecycle & Trust Signals Open/close events, days open, repost detection, stale and ghost scoring, apply-channel type	posting / event	47,848 close events 8,281 reopen events 43,676 scored postings	daily	events since 2026-07-23	included in the postings sample (status, closed at, days open, repost_count, trust_level, lifecycle_events)

Tech Salary Benchmarks Percentiles of stated and estimated pay by role, seniority and geography	benchmark cell	37,195 live postings with stated pay 62,592 with model estimate	daily	rolling window (live + 180 days of closed postings); monthly snapshots on request	CSV, 1,214 cells
Taxonomies Delivered with every dataset	reference	3,428 technologies 1,086 roles in 20 families 651 industry markets 17,679 cities, 236 countries	monthly	versioned	on request

3. Coverage snapshot: postings

121,355 live postings	7,855 employers hiring now	13,398 new in the last 7 days
48,653 new in the last 30 days	34,620 closed in the last 30 days	133 countries with live postings
283 cities with 50+ live postings		

Field coverage on live postings

Field	Postings	Share
Role mapped to taxonomy	120,673	99.4%
Seniority resolved	100,356	82.7%
Technologies tagged (860,756 tags, 2,123 distinct)	105,081	86.6%
Stated salary, normalized to annual USD	37,195	30.6%
Model salary estimate (postings without stated pay)	62,592	51.6%
Remote	24,399	20.1%
Hybrid	26,427	21.8%
On-site	70,515	58.1%
Visa sponsorship stated	4,853	4%
Equity mentioned	14,553	12%

Top countries (office location of live postings)

Country	Live postings
United States	38,944
India	13,869
Poland	6,428
United Kingdom	5,218
Canada	3,391
France	2,683
Germany	2,611
Singapore	2,065
Spain	1,432
Australia	1,266
Malaysia	1,130
Mexico	1,076
Portugal	1,029
Brazil	1,028
China	985

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

Trust scored (stale/ghost/ok)	37,527	30.9%
----------------------------------	--------	-------

United Kingdom, Canada, EU, Singapore.

Source type

Source	Live postings	Share
Employer ATS boards and APIs	107,634	88.7%
Employer career pages (schema.org / HTML)	9,085	7.5%
Partner feeds and APIs	4,636	3.8%

Role families

Family	Live postings
Backend	29,497
Product	12,027
DevOps	10,290
Analytics	9,449
AI/ML	8,600
Data Science	7,689
QA	6,619
Hardware	6,160
Security	5,685
Executive	4,112
Design	3,439
Management	3,277
Enterprise Apps	2,977
Support	2,898
Frontend	2,099

Seniority

Level	Live postings
Senior	42,508
Staff +	12,408
Lead	14,382
Head / Executive	6,358
Middle	14,402
Junior	9,054
Intern	1,244
Not resolved	20,999

Employment type

Type	Live postings
Full-time	83,937
Contractor	4,576
Internship	4,635
Part-time	642

ATS platforms (employers with live postings)

ATS	Employers	Live postings
Workday	578	31,149
Greenhouse	1,123	21,352
SmartRecruiters	401	16,054
Ashby	1,307	14,347
Lever	619	10,411
Workable	772	7,058
Oracle	86	3,126
Eightfold	13	1,388
Trafit	54	1,320
Recruitify	14	850

Mobile	1,693
Networking	1,182
Robotics	966
Marketing	807
Games	629

Freelance	421
Co-founder	247
Not stated	26,897

Work at a Startup	215	618
Imagine	1	606
Recruitee	82	582
eRecruiter	72	553
Schena.org career pages	39	430
BamborHR	97	369

60+ ATS platforms supported in total.

4. Coverage snapshot: companies

Attribute	Companies	Note
Company records	93,957	technology employers and companies with technology hiring
Hiring now (live postings)	7,855	hiring signals are computed for these
Industry classified	43,635	651-market taxonomy, two levels
Description	61,539	from company website
Office locations	27,017	HQ flag, city and country
Founded year	25,666	
ATS platform identified	8,465	
Technology stack from hiring (385,497 company-technology rows)	11,753	first-seen / last-seen dates per technology
Weekly hiring snapshots since 2026-07-20	18,374	74,810 weekly rows
Hiring signal scorecards (demand, pay, trust, spikes)	10,458	

Dated growth events (geo expansion, tech added/removed, hiring spike, funding)	287,085	events, not companies
Social profiles (LinkedIn, GitHub, X, company-level)	41,184	
Size band (evidence-based estimate)	3,738	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	1,358	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 (47,848 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 43,676 postings carry a score (16,011 ok, 14,895 stale, 12,770 ghost); coverage grows as the corpus accumulates observation history.
- **History.** Closed postings are retained (38,318 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.

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.

- Free sample files (this catalog) and a trial period with full fields on a limited slice.

- Rate-limited collection; source URL and timestamps retained per record for audit.

8. Sample files

File	Rows	Content
alio_sample_job_postings_20260904.csv	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.
alio_sample_company_hiring_signals_20260904.csv	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.
alio_sample_salary_benchmarks_20260904.csv	1,214	Benchmark cells (role and role family x seniority x country / Remote / All) computed from 98,910 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 2026-09-04 and exclude regional job board data. Alion, alion.io.