

Alion Data Dictionary

Field definitions for the three sample datasets, with fill rate and an example value measured on the sample files published on 2026-09-21. The full products carry the same fields; the postings product also delivers the full description text.

Conventions: timestamps are UTC ISO 8601; dates are YYYY-MM-DD; money fields are integers in the stated currency or in USD as named; list fields are pipe-separated (|); counts inside lists are in parentheses; booleans are 0/1; empty means not available for the record.

Job Postings (with lifecycle, trust and salary fields)

File: [alio_sample_job_postings_20260921.csv](#). Unit: one row = one job posting. Rows in sample: 2,500.

Field	Type	Description	Fill	Example
job_id	integer	Stable Alion identifier of the posting.	100%	254320
alio_url	url	Canonical Alion page of the posting.	100%	https://alio.io/job/zippi-senior-software-engineer
title	text	Job title as published by the employer.	100%	Senior Software Engineer
role	text	Normalized role from the Alion role taxonomy (1,086 roles).	100%	Backend
role_family	text	Top level of the role taxonomy (20 families: Backend, AI/ML, DevOps, Data Science, Security...).	100%	Backend
seniority	enum	Resolved seniority level: Intern, Junior, Middle, Senior, Staff+, Lead, Head / Executive. Empty when the posting gives no usable signal.	62%	Senior
employment_type	enum	Full-time, Part-time, Contractor, Co-founder, Freelance, Internship. Empty when not stated.	75%	Full-time
work_mode	enum	On-site, Hybrid, Remote, Unspecified.	100%	Remote
remote_scope	enum	For remote postings: remote within stated countries, remote within stated regions (EU, LATAM, EMEA...), remote worldwide, or scope unresolved. Empty for on-site and hybrid postings.	20%	Remote within stated countries
hiring_regions	list	Regions or countries the employer hires from, pipe-separated (e.g. European Union, United States).	2%	Europe
hiring_geo_confidence	enum	How the hiring geography was established: explicit (stated in text), structured (from structured location data), inferred.	100%	structured
locations	list	Office locations resolved to cities (and regions), pipe-separated.	75%	São Paulo

countries	list	Countries of the office locations, pipe-separated.	75%	Brazil
company_id	integer	Alion identifier of the employer. Joins to the company dataset.	100%	9277
company_name	text	Employer name.	100%	Zippi
company_domain	text	Employer primary web domain.	100%	zippi.com.br
company_alion_url	url	Alion company page.	100%	https://alio.io/company/zippi
company_industry	text	Primary industry of the employer (top level of 651-market taxonomy).	76%	Financial Services
company_industry_segment	text	Industry segment (second level).	76%	FinTech
company_size	enum	Headcount band: 1-10, 11-50, 51-200, 201-500, 501-1000, 1001-5000, 5000+. Evidence-based estimate; empty when confidence is low.	68%	51-200
company_is_agency	boolean	1 when the poster is a staffing agency or recruitment intermediary rather than the hiring employer.	100%	0
ats_vendor	text	Applicant tracking system the employer publishes on (Workday, Greenhouse, Ashby, Lever, SmartRecruiters, Workable...).	95%	Work at a Startup
salary_stated_min	integer	Lower bound of the compensation stated in the posting, in the stated currency and period.	29%	50000
salary_stated_max	integer	Upper bound of the stated compensation.	28%	120000
salary_currency	text	ISO code of the stated currency (48 currencies supported).	30%	USD
salary_period	enum	Period of the stated figures: year, month, week, day, hour.	30%	year
salary_stated_usd_annual	integer	Stated compensation normalized to annual USD (midpoint of the range, monthly/hourly figures annualized).	30%	120000
salary_estimate_min_usd	integer	Model-based estimate, lower bound, annual USD. Produced for postings without a usable stated salary.	53%	103000
salary_estimate_max_usd	integer	Model-based estimate, upper bound, annual USD.	53%	213000
salary_estimate_method	enum	Which benchmark cell produced the estimate (role+seniority+country+remote; role+seniority+country; global role scaled by country factor; ...). Lower specificity = wider band.	52%	global role cell scaled by country factor
salary_estimate_sample_n	integer	Number of stated-salary observations behind the benchmark cell.	53%	2106

has_equity	boolean	1 when equity or stock compensation is mentioned.	100%	⁰
visa_sponsorship	boolean	1 when the posting states visa sponsorship is available.	100%	⁰
relocation_package	boolean	1 when relocation support is stated.	100%	⁰
experience_years_min	integer	Minimum years of experience stated in the posting.	44%	³
technologies	list	Technologies, tools and frameworks extracted from the posting, mapped to the 3,428-entry technology taxonomy, pipe-separated, required first. Capped at 25 in the sample.	59%	HIPAA
technology_count	integer	Number of distinct technologies tagged.	100%	⁰
benefits	list	Benefits mentioned, mapped to a 56-entry taxonomy, pipe-separated.	43%	Equity Parental leave Stock options
apply_channel	enum	direct_ats (application form on the employer ATS), external_link, email, redirect (posting redirects to another posting), unknown.	100%	direct_ats
apply_url	url	Direct application URL (employer ATS or career page).	96%	https://join.com/companies/conni-partner/16724563
source_type	enum	employer ATS board, employer career page, partner feed / api.	100%	employer ATS board
source_url	url	URL the record was collected from. Retained for every record for provenance.	100%	https://www.workatastartup.com/jobs/68035
first_seen_at	datetime (UTC)	When Alion first collected the posting.	100%	2026-09-05T11:15:29Z
employer_posted_date	date	Publication date reported by the employer ATS, when available.	94%	2026-09-19
last_verified_at	datetime (UTC)	Last time the posting was confirmed on its source board (for closed postings: the close time).	100%	2026-09-21T08:17:07Z
status	enum	live or closed. The sample contains 2,200 live and 300 recently closed postings.	100%	live
closed_at	date	Date the posting disappeared from its source (closed postings only).	12%	2026-09-21
days_open	integer	Days between first_seen_at and now (live) or closed_at (closed).	100%	¹⁶
repost_count	integer	How many times the same vacancy was re-published by the employer. Present for scored postings.	39%	¹

trust_level	enum	ok, stale, ghost, or not_scored. Stale = open far longer than typical time-to-fill for the role; ghost = repeated reposts or company-wide stale pattern. Scored for postings with enough observation history (about 31% of live postings today, growing).	39%	ok
trust_flags	list	Reasons behind the trust level: stale, repost, company_stale.	35%	stale
lifecycle_events	list	Event log of the posting: open@date, close@date, reopen@date, expire@date, pipe-separated.	100%	open@2026-09-05
description_chars	integer	Length of the full description text (full text delivered in the product; the sample carries an excerpt).	100%	3736
description_excerpt	text	First 400 characters of the plain-text description.	100%	

Company Hiring Signals and Firmographics

File: [alion_sample_company_hiring_signals_20260921.csv](#). Unit: one row = one employer. Rows in sample: 1,500.

Field	Type	Description	Fill	Example
company_id	integer	Alion identifier of the company. Joins to the postings dataset.	100%	5239
company_name	text	Company name.	100%	Capital One
domain	text	Primary web domain.	100%	capitalone.com
alion_url	url	Alion company page.	100%	https://alio.io/company/capitalone
industry	text	Primary industry (top level).	86%	Financial Services
industry_segment	text	Industry segment (second level).	86%	Financial Services
size_band	enum	Headcount band (1-10 ... 5000+), evidence-based estimate.	84%	5000+
size_confidence	decimal	Confidence of the size estimate, 0-1.	82%	0.68
founded_year	integer	Founding year.	67%	1994
hq_city	text	Headquarters city.	99%	McLean
hq_country	text	Headquarters country.	99%	United States
offices_on_file	integer	Number of office locations on file.	100%	54
is_staffing_agency	boolean	1 for staffing agencies and recruitment intermediaries.	100%	0
ats_vendor	text	Applicant tracking system in use.	98%	Workday
ats_board_last_verified_at	datetime (UTC)	Last time the company job board was fully fetched and reconciled.	99%	2026-09-21T12:14:00Z

open_jobs	integer	Live postings right now.	100%	1397
new_jobs_30d	integer	Postings first seen in the last 30 days.	100%	733
new_jobs_90d	integer	Postings first seen in the last 90 days.	100%	1261
new_jobs_prev_30d	integer	Postings first seen 30-60 days ago (baseline for velocity).	100%	559
closed_jobs_30d	integer	Postings that disappeared from the source in the last 30 days.	100%	377
hiring_velocity_30d_ratio	decimal	new_jobs_30d / new_jobs_prev_30d. Above 1 = accelerating, below 1 = slowing, "new" = no baseline.	95%	1.31
open_jobs_weekly_last_9w	list	Weekly time series of live postings, oldest first, pipe-separated (weekly snapshots since July 2026).	100%	4157 59 59 46 51 51 1525 1392
last_job_posted	date	Date of the most recent posting.	83%	2026-09-21
oldest_open_job_days	integer	Age in days of the oldest live posting.	100%	426
open_senior_plus	integer	Live postings at Senior level or above.	100%	1209
open_remote	integer	Live remote postings.	100%	9
open_hybrid	integer	Live hybrid postings.	100%	174
open_visa_sponsorship	integer	Live postings offering visa sponsorship.	100%	0
open_with_equity	integer	Live postings mentioning equity.	100%	1
top_roles	list	Most hired roles with counts, e.g. Backend(12) AI/ML(5).	100%	Backend(836) Frontend(166) Mobile(32) QA(16) DevOps(9)
hiring_countries	list	Countries of live postings with counts.	97%	United States(1354) Australia(17) Poland(2) Luxembourg(1)
hiring_cities	list	Cities of live postings with counts (postings located at country level only are not counted here).	96%	Singapore(793) Nanyang(21) Stamford(1)
technologies_in_stack	integer	Distinct technologies observed in the company postings over time.	100%	355
top_technologies	list	Most demanded technologies with posting counts.	96%	Microsoft Office(7) Outlook(4) Microsoft Project(2)
technologies_first_seen_90d	list	Technologies that appeared in the company hiring for the first time in the last 90 days, with the first-seen date. Technology adoption signal.	95%	Microsoft Office(2026-09-18) WhatsApp(2026-09-10)
stated_salary_n	integer	Live postings with a stated salary.	100%	1244

avg_stated_salary_usd	integer	Average stated salary of live postings, annual USD.	75%	196337
avg_estimated_salary_usd	integer	Average model-estimated salary of live postings, annual USD.	93%	128920
repost_count	integer	Reposted vacancies detected for the company.	100%	64
ghost_postings	integer	Live postings scored as ghost.	100%	0
max_days_open	integer	Longest-open live posting, days.	100%	426
hiring_spikes	integer	Detected hiring spikes (week-over-week jumps).	100%	2
has_careers_page	boolean	Careers page detected on the company website.	100%	1
has_hiring_banner	boolean	"We are hiring" banner detected on the website.	100%	0
demand_score	integer 0-100	Composite hiring demand score.	100%	91
pay_score	integer 0-100	Composite pay level score relative to market.	100%	59
hiring_signal_score	integer 0-100	Overall hiring signal score.	100%	74
hiring_signal_level	enum	Letter grade of the hiring signal score (A-D).	100%	A
last_funding_round	text	Latest funding round type on file (Seed, Series A...). Sparse: funding is tracked for a subset of venture-backed companies.	3%	IPO
last_round_amount_usd	integer	Amount of the latest round, USD.	3%	110000000
last_round_year	integer	Year of the latest round.	3%	2004
funding_rounds_count	integer	Rounds on file.	100%	0
total_raised_usd	integer	Sum of rounds on file.	3%	110000000
revenue_year	integer	Fiscal year of the latest revenue figure (public filings). Sparse.	9%	2025
revenue_min_usd	integer	Revenue lower bound, USD.	9%	8062000000
revenue_max_usd	integer	Revenue upper bound, USD.	9%	8062000000

growth_events_90d	list	Detected events in the last 90 days with counts: geo_expansion, tech_added, tech_removed, hiring_spike, funding.	100%	tech_added(467) 20(1076) 25(9) 26(11) 27(1070)
linkedin_url	url	Company LinkedIn page (company-level, public).	76%	https://www.linkedin.com/company/capital-one
github_url	url	Company GitHub organization.	3%	https://github.com/salesforce
x_url	url	Company X (Twitter) account.	48%	https://x.com/askcapitalone
description_excerpt	text	First 240 characters of the company description.	97%	Welcome to Tesla Laboratories. Cyber Security.

Tech Salary Benchmarks

File: [alio_sample_salary_benchmarks_20260921.csv](#). Unit: one row = one benchmark cell (role or role family x seniority x geography). Rows in sample: 4,094.

Field	Type	Description	Fill	Example
segment_level	enum	role (specific role) or role_family (top taxonomy level).	100%	role
role_family	text	Role family.	100%	AI/ML
role	text	Specific role (empty for role_family rows).	50%	AI/ML
seniority	enum	Intern, Junior, Middle, Senior, Staff+, Lead, Head / Executive.	100%	Junior
geography	text	Country of the office location, "Remote" for remote postings, or "All countries".	100%	Singapore
stated_n	integer	Number of postings with a stated salary in the cell.	100%	9
stated_p10_usd	integer	10th percentile of stated salaries, annual USD. Published when stated_n >= 8.	62%	250000
stated_p25_usd	integer	25th percentile.	62%	250000
stated_p50_usd	integer	Median.	62%	250000
stated_p75_usd	integer	75th percentile.	62%	250000
stated_p90_usd	integer	90th percentile.	62%	250000
estimated_n	integer	Number of postings with a model estimate in the cell.	100%	32
estimated_p25_usd	integer	25th percentile of estimate midpoints. Published when estimated_n >= 30.	85%	148500
estimated_p50_usd	integer	Median of estimate midpoints.	85%	153250

estimated_p75_usd	integer	75th percentile of estimate midpoints.	85%	167000
currency	text	USD.	100%	USD
period	text	annual.	100%	annual
observation_window	text	Postings included: live, plus postings closed within the last 180 days.	100%	live + closed within 180 days
as_of	date	Computation date.	100%	2026-09-21

Enumerations

Field	Values
seniority	Intern, Junior, Middle, Senior, Staff+, Lead, Head / Executive
employment_type	Full-time, Part-time, Contractor, Co-founder, Freelance, Internship
work_mode	On-site, Hybrid, Remote, Unspecified
remote_scope	Remote within stated countries; Remote within stated regions; Remote worldwide; Remote, scope unresolved
hiring_geo_confidence	explicit, structured, inferred
apply_channel	direct_ats, external_link, email, redirect, unknown
source_type	employer ATS board; employer career page; partner feed / api
status	live, closed
trust_level	ok, stale, ghost, not_scored
trust_flags	stale, repost, company_stale
lifecycle_events	open, close, reopen, expire (each with a date)
salary_estimate_method	role+seniority+country+remote cell; role+seniority+country cell; global role cell scaled by country factor; global role+seniority cell; role+country, seniority unknown; seniority bridged from stated years
growth_events_90d	geo_expansion, tech_added, tech_removed, hiring_spike, funding
hiring_signal_level	A, B, C, D
size_band / company_size	1-10, 11-50, 51-200, 201-500, 501-1000, 1001-5000, 5000+
last_funding_round	Pre-seed, Seed, Angel, Series A-H, Venture (series unknown), Private equity, Debt financing, Grant, Equity crowdfunding, Convertible note, Secondary market, Post-IPO equity, Post-IPO debt, Acquired, IPO, Direct listing, SPAC, Other

Taxonomies delivered with the data

Taxonomy	Entries	Structure
Technologies	3,549	tree with parent links (e.g. PyTorch → Deep learning frameworks → AI/ML), aliases and detection patterns

Roles	1,198	21 families, specific roles beneath, multilingual titles
Industries	709	two levels: industry and segment
Locations	22,579 cities, 236 countries	city → region → country, with regions (EU, EEA, LATAM, MENA...) for hiring scope
Currencies	48	ISO codes with daily USD rates used for normalization
Benefits	56	flat list

Alion, alion.io. Dictionary version 2026-09-21.