PITON Global

EXECUTIVE WHITE PAPER · WP-44

DATA SERVICES · PHILIPPINES

The data-fitness standard: the executive economics of data services outsourcing to the Philippines.

An analysis of why records processed is a volume vanity metric, how data fitness and lifecycle integrity — never processing throughput — decide the true cost of a data operation once downstream errors, duplicate and stale records, broken pipelines, and rework are counted, and the vendor-selection discipline that delivers data fit for the use it was built for.

AUTHORS

John Maczynski — Chief Executive Officer
Ralf Ellspermann — Chief Strategy Officer

JULY 2026
MANILA · ESTABLISHED 2001

Contents

ABOUT THIS SERIES — VOLUME 44

Every data-services proposal is priced per record processed, because volume is easy to count. This volume is about the only data outcome a business actually depends on: the record captured accurately, kept clean across its lifecycle, and delivered fit for the use it was built for. The gap between processing a record and making it fit-for-use is where downstream errors and rework live. The full series lives at piton-global.com.

THE PITON-GLOBAL LEADERSHIP TEAM

About the authors

John Maczynski

Chief Executive Officer

John has bought data-services capacity for three decades, and learned that the fast operation that processes every record and gets the data wrong is the most expensive one a business can hire — because a bad record is a wrong decision downstream, a duplicate inflates every count built on it, and a broken pipeline stops the analytics that depend on it. His standing question ignores the records-processed report: of the records you processed, how many were accurate, clean, and fit for the use they were built for? Applied at PITON-Global, that question is what clients get on every data-services sourcing decision — personally, with no account layer in between.

j.maczynski@piton-global.com
+1 402-598-8740

Ralf Ellspermann

Chief Strategy Officer

Ralf has spent twenty-five years running Philippine data-operations floors, and he judges a data team by its data-fitness rate and lifecycle integrity, not its records-per-hour: whether the data was captured right and stayed clean end to end, or was processed fast and broke downstream. The data contract in Section 3 codifies what he learned rebuilding teams that processed records fast and shipped dirty data. He now reads data-services providers the way he once ran them — from the downstream-error log and the data-quality audit backward.

r.ellspermann@piton-global.com
Manila, Philippines

PITON Global

ABOUT PITON-GLOBAL

PITON-Global is a buyer-side advisory practice, Manila-based since 2001, that has never operated a delivery floor or taken a placement commission. Its stock-in-trade is knowing the Philippine provider market at the operating level — 1,000+ mapped operations, re-verified continuously, including each data team's real data-fitness rate and lifecycle integrity rather than its records-per-hour headline — and

converting that knowledge into a shortlist of six to ten genuinely qualified partners per engagement. Buyers from Tripadvisor, Garmin, TSYS, HealthPartners, Yetipay, Workrise, and Norman Disney & Young have run their searches through the practice.

SECTION 01

Executive summary

-59%
Fully-loaded cost per fit-for-use record vs. US onshore

99.5%+
Data-accuracy rate on vetted teams, up from 92–96%

-66%
Reduction in downstream errors and rework on vetted teams

6.4×
First-year ROI, reconstructed case (Section 6)

A data-services engagement is bought on the price per record and paid for on the data it gets wrong. The vendor reports tens of millions of records processed at a fraction of the onshore cost; the business finds the errors that surfaced three systems downstream, the duplicates inflating every count and campaign built on them, the stale records nobody refreshed, and the pipeline that broke silently and starved the analytics. The operation hit its records-processed number and the business inherited dirty data, broken reporting, and a rework bill that dwarfs the per-record saving. The gap between processing a record and making it fit-for-use is where the real cost of cheap data work hides, because a wrong or dirty record is not throughput delivered — it is a wrong decision waiting downstream.

The Philippines — English-fluent, detail-disciplined, with a deep data-operations talent pool across entry, cleansing, enrichment, and pipeline work — is where data services is delivered with the most data-fitness discipline, because that pool can run the accurate capture, lifecycle hygiene, and fit-for-use validation the work demands. Five findings:

1. **Records processed is the industry's most flattering number.** Any team can push records through at volume. Vetted teams are measured on outcome: data accuracy, lifecycle integrity, and fitness for use. Weak providers quote a low per-record price the business repays in downstream errors, rework, and broken analytics. Section 2 shows the gap.

2. **The fit-for-use record, not the fast keystroke, is the product.** Data a business can rely on comes from capturing it accurately, keeping it clean across its lifecycle, and validating it against the use it serves — not from clearing a record queue. Teams that deliver fitness feed clean decisions; teams that process loosely ship dirty data and rework that costs more than the records.

3 **A bad record cuts both ways — error and decay.** A wrong value corrupts the decision it feeds; a duplicate or stale record quietly decays every count and model built on it. Vetted teams hold data accuracy above 99.5% precisely because both tails cost — the error that misleads and the decay that accumulates unseen.

4 **The downstream error and the rework, not the keystroke, are what cost.** A record done wrong costs the bad decision it feeds, the pipeline it breaks, and the rework to find and fix it — many times the per-record saving. Teams that deliver fitness cut downstream errors and keep pipelines whole, and that protected data — not the offshore rate alone — is where the case study's return is built.

5 **Contract on the fit-for-use record, not the processed one.** The contracting failure of data services is paying per record and hoping the data is accurate and fit. Vetted deals price against data-accuracy floors, lifecycle-integrity targets, and fitness-for-use minimums, making usable data the team's problem, which is where it belongs.

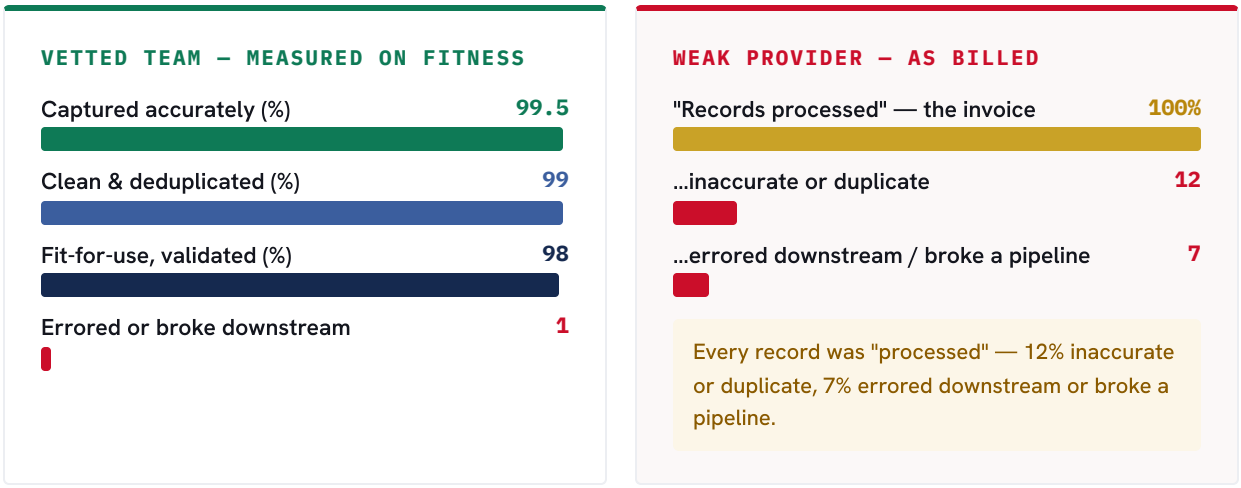
The intended reader is a chief data officer, head of data operations, or COO buying data services per record and paying for it in downstream errors, rework, and broken analytics. Sections 2–4 separate the teams that deliver data fitness from the teams that process records; Sections 5–7, the method and the proof.

SECTION 02

The volume mirage: records processed versus fit-for-use data

Figure 1 shows one month of a business's data pipeline two ways: as the records-processed report a weak vendor bills, and as the data-fitness reality the downstream systems, the analytics, and the audit experience. The records-processed number is identical in spirit to a chatbot's containment rate — a figure engineered to look like output while errors, duplicates, and broken pipelines accumulate underneath.

FIG. 1 — ONE MONTH OF THE PIPELINE: DATA-FITNESS QUALITY VS. REPORTED VOLUME



Left: PITON-Global engagement instrumentation, 2025–26, data fitness audited. Right: composite of pre-engagement vendor reporting reviewed in diligence.

The economics follow the data fitness, not the records count. On the left, ninety-nine-and-a-half percent of records are captured accurately and ninety-eight percent are validated fit-for-use, so downstream systems run clean and analytics can be trusted; on the right, a cheap per-record price ships twelve percent inaccurate or duplicate and errors seven downstream, paying in wrong decisions, broken pipelines, and rework at once. A low per-record price that ships dirty data is not a saving; it is downstream error and decay, delayed and compounding. The data-fitness audit — accuracy, dedup and lifecycle integrity, fitness for use — is the cheapest due-diligence instrument in this paper.

SECTION 03

The data contract: capture it accurately, keep it clean, deliver it fit-for-use

Figure 2 lays out the three clauses of a working data-services contract — the machinery a floor read verifies before any per-record price is believed.

FIG. 2 — THE DATA CONTRACT, AS OPERATED ON VETTED TEAMS

CAPTURE IT ACCURATELY

Every field is entered or ingested against the source with a double-key or validation check on the values that matter, to the schema and the business rules — and an ambiguous or malformed input flagged rather than guessed. The data is made correct at the point of capture, so nothing dirty enters the pipeline.

KEEP IT CLEAN

Records are deduplicated, standardized, enriched, and refreshed on a cadence, with lineage tracked and stale data retired — so the dataset stays clean across its whole lifecycle rather than decaying between projects. Clean-over-time is what keeps every count and model built on the data honest.

DELIVER IT FIT-FOR-USE

Data is validated against the use it serves — the report, the model, the system it feeds — and delivered on-schema, on-time, with pipelines monitored so a break is caught before it starves anything, and a data-quality report the business can read. Fit-for-use is the only definition of done that the downstream actually feels.

↓ yields ↓

THE FIT-FOR-USE RECORD — 99.5%+ ACCURACY · CLEAN LIFECYCLE · FIT-FOR-USE

Records are captured accurately, kept clean across the lifecycle, and delivered fit for the use they serve — so downstream errors disappear, pipelines stay whole, and analytics can be trusted. The operation gets cheaper per fit-for-use record and more reliable at the same time, which is the entire promise of outsourcing data services to people who measure data fitness, not records processed.

Operating parameters from PITON-Global data-operations engagements, 2025–26.

In diligence, the contract is tested from the downstream's side: pull processed records and re-validate a sample against the source and the schema for true accuracy; audit for duplicates and staleness; trace a pipeline end-to-end for breaks. Perhaps one team in ten survives it cleanly. Steps 3 and 5 of the framework exist to find that one before the contract does — because everything in Table 3 of Section 6 depends on data fitness, not records processed.

SECTION 04

Cost and performance: cost-per-fit-for-

use-record economics and benchmarks

TABLE 1 — FULLY-LOADED COST PER FIT-FOR-USE RECORD: THREE OPERATING MODELS (2026)

OPERATING MODEL	COST / 1K RECORDS	VS. US BASELINE
US onshore in-house data team	$34.00	baseline
Low-cost offshore data mill (per-record)	$23.00	–32%
Vetted Manila team — accurate, clean, fit	$13.90	–59%

Basis: PITON-Global engagement pricing 2025–26; cost per fit-for-use record = fully-loaded operating cost divided by records delivered accurate, clean, and validated fit-for-use (inaccurate, duplicate, and errored records excluded), per 1,000 records; the data mill looks cheap per record but downstream errors, broken pipelines, and rework lift its true cost per fit record; US comparator fully-loaded in-house cost.

TABLE 2 — DATA-SERVICES BENCHMARKS, VETTED VS. BASELINE (2025–26)

METRIC	VETTED TOP TIER	BASELINE	EXECUTIVE READING
Data-accuracy rate	99.5%+	92–96%	The only data number decisions follow
Duplicate / staleness rate	<1%	5–12%	Clean, or quietly decaying
Downstream errors vs. baseline	–66%	baseline	"Deliver it fit" — held or not
Pipeline uptime / on-schema delivery	99.5%+	95–98%	Analytics fed, or starved
Rework rate vs. baseline	–63%	baseline	Done once, or found and refixed

Source: PITON-Global data-operations engagement data 2025–26; baseline = typical pre-engagement operations billed on records processed.

The strategic read: the data mill's –32% per record is a false economy — downstream errors, broken pipelines, and rework lift its true cost per fit record far above the per-record gap, and a wrong decision made on dirty data is a cost no rate recovers. The vetted Manila team's –59% is real because it is measured where the value is: the record captured accurately, kept clean, and delivered fit for use. Buy on records processed and you buy the right column of Figure 1; buy on cost-per-fit-for-use-record and you buy data fitness — the same discipline every volume in this series keeps arriving at from a different door.

SECTION 05

Sourcing discipline: the 7-Step Vendor Vetting Framework, applied to data services

Data-services diligence adds one hard requirement to every step: a per-record price is real only when the data fitness behind it survives a re-validation against the source and the schema. The funnel below is representative of a 30–70-seat data-operations search.

FIG. 3 — THE 7-STEP FRAMEWORK AS A DATA-SERVICES FUNNEL

	STEP	WHAT IT ELIMINATES IN A DATA-SERVICES SEARCH	FIELD REMAINING
1	**Requirements definition**	Throughput-first scoping: fixes the data-accuracy floor, the lifecycle-integrity target, the pipeline-uptime SLA, and the fitness-for-use standard the operation must meet by data domain	1,000+
2	**Market mapping**	Data mills: operations that bill record volume and treat dirty data and broken pipelines as the client's problem, and generalist floors with no data-quality or pipeline depth	~46
3	**Capability screening**	Teams without the machinery: audited accuracy history, double-key/validation process, dedup-and-cleansing workflow, pipeline monitoring, and coverage across your data domains	~13
4	**Compliance & security audit**	Weak governance: PII protection and data residency, access controls on the data platform, lineage and audit trail, breach discipline, SOC 2 scope covering the data stack	~8
5	**Operational diligence**	The re-validation: re-validate processed records against the source and schema for true accuracy; audit for duplicates and staleness; trace a pipeline for breaks; interview on how they handle a malformed feed	4–6
6	**Competitive RFP & negotiation**	Per-record pricing: finalists bid against data-accuracy floors, lifecycle-integrity targets, and pipeline-uptime SLAs — all carrying credits for records that come back inaccurate or break downstream	2–3
7	**Transition & launch governance**	Cold cutover: the team calibrates against your schema, data domains, and platform on a sample before it owns production, with the accuracy floor certified before steady state, PITON-Global embedded buyer-side	1

Field-remaining counts represent a 30–70-seat data-operations search, PITON-Global engagements 2024–26.

The standing guarantees hold: **vendor neutrality** — PITON-Global operates no data floors, processes no records, and takes no placement economics, so the shortlist reflects audited data-fitness performance, not inventory; **right-sizing** — teams for whom your operation is a flagship account. Six to ten genuinely qualified providers, delivered within 48–72 hours, free to the buyer, competing through a fully transparent, fair, comprehensive, and highly competitive RFP.

SECTION 06 · ANONYMIZED CASE STUDY

Re-basing the operation on data fitness: a 50-seat data rebuild, reconstructed

Engagement DS-057 is a US enterprise — processing roughly 240,000,000 records a year across data entry, cleansing, enrichment, and pipeline operations — that had offshored data services to a per-record mill. The record price was excellent; the data was not: data accuracy sat near 94%, duplicates and stale records were inflating counts, pipelines broke silently, and downstream teams no longer trusted the numbers. The chief data officer wanted the scale without the dirty data and the rework. Identifying details are anonymized under NDA.

SELECTION (WEEKS 0–8, RE-VALIDATIONS)

Requirements fixed a 99.5% data-accuracy floor, a sub-1% duplicate/staleness target, and a 99.5% pipeline-uptime SLA, with a mandatory double-key validation and dedup workflow. Mapping produced 46 claimants; the machinery screen — audited accuracy, validation process, dedup/cleansing, pipeline monitoring — left 13; the security-and-governance audit left 8. Blind re-validations of each finalist's processed records against source and schema shortlisted 5. A Manila team won priced against data accuracy and pipeline uptime, with a re-validation written into the reporting spec.

TRANSITION (MONTHS 2–6, CALIBRATED FIRST)

The team calibrated against the enterprise's schema, data domains, and platform on sample loads and cleaned the existing duplicate/stale backlog before owning production; the accuracy floor was certified in month 3. The validation-and-monitoring loop fed feedback from day one. By month 6 data accuracy reached 99.6%, duplicate/staleness fell below 1%, downstream errors dropped 66%, and pipeline uptime hit 99.5%. PITON-Global chaired the data-fitness review board to steady state.

TABLE 3 — DS-057 FIRST-YEAR VALUE BUILD-UP (US$)		
VALUE COMPONENT	YEAR 1	BASIS
Labor arbitrage on the 50-seat data operation	$1.52M	93.8k productive hrs × ($26.10 − $9.90)
Downstream-error & rework cost removed	$0.78M	Error-hunting and refix cost avoided
Decision-quality value (trusted analytics)	$0.42M	Bad-decision and mis-targeting cost avoided
Pipeline-downtime cost avoided	$0.18M	Analytics-starvation and outage cost removed
Gross value created	**$2.90M**	
Less: engagement & transition costs	($0.45M)	Calibration, backlog cleanup, schema/platform integration; advisory fee $0 (provider-paid model)
Net value ÷ cost = first-year ROI	6.4×	$2.90M ÷ $0.45M

Figures rounded; accuracy, dedup, and pipeline effects audit-verified. Method in Methodology & notes.

Fifty-two percent arbitrage, forty-eight percent from removed downstream errors and rework, trusted analytics, and pipeline uptime — a typical split for the series, earned because the operation was re-based on data fitness rather than record volume. Every line traces to a diligence step: the accuracy floor to Step 1, the fitness-based pricing to Step 6, the validation workflow to Step 3's contract. The chief data officer's line at the year-one review: "The record price barely moved. What moved is that the numbers are trusted again — nothing breaks downstream, and we stopped hunting for errors."

SECTION 07

The executive playbook: governance for the first 180 days

A data-services transition succeeds when the team is measured and paid on data accuracy and lifecycle integrity — client in command throughout.

DAYS 0–30

Define fitness, contract on it

Set the data-accuracy floor, the lifecycle-integrity target, and the pipeline-uptime and fitness-for-use thresholds with data leadership before launch, and document the schema, validation, and dedup workflow by data domain. Contract against data

	accuracy and lifecycle integrity — never per record alone — with floors and SLAs carrying credits. Stand up the re-validation as the program's arbiter of truth.
DAYS 30–90 **Calibrate, clean the backlog**	Calibrate the team against your schema, data domains, and platform and clean the existing duplicate/stale backlog before it owns production; certify the accuracy floor and dedup loop. Switch on pipeline monitoring from day one. Publish the data dashboard: accuracy, duplicate/staleness, downstream errors, pipeline uptime, rework rate.
DAYS 90–180 **Own production, certify steady state**	Hand production to the team only after it holds the accuracy floor and pipeline-uptime SLA. Review the data board monthly with the downstream-error and staleness trends on the table. Certify steady state on two months at floor including a schema-change cycle; hand over on documented criteria; keep the quarterly re-validation standing.

THE FIVE FAILURE MODES WE ARE HIRED TO PREVENT

Paying per record while data ships dirty · processing records instead of capturing them accurately · letting duplicates and stale records decay the dataset · running pipelines with no monitoring until one breaks silently · owning production before data fitness is proven. All five are settled at selection and contract.

The data-services argument, in one sentence: a processed record is activity and a fit-for-use record is a decision fed clean, and a vetted Philippine team delivers −59% cost per fit-for-use record with 99.5%+ data accuracy, sub-1% duplication, and downstream errors cut by two-thirds — for buyers who re-validate the data fitness instead of counting records processed, and compete the finalists through a fully transparent, fair, comprehensive, and highly competitive RFP. Anyone can process a record. We make sure you buy the team whose data is fit for the use you built it for.

METHODOLOGY & NOTES

How the numbers were built

Accuracy and fitness figures (Fig. 1, Table 2) derive from engagement instrumentation in PITON-Global data-operations engagements, 2025–26. Data accuracy and lifecycle integrity are measured by re-validating processed records against the source and schema and auditing for duplicates and staleness; downstream errors, pipeline uptime, and rework are measured against the client's pre-engagement baseline. The weak-provider column is a composite of pre-engagement vendor reporting reviewed in diligence, shown for contrast.

Cost-per-fit-for-use-record figures (Table 1) are fully loaded — wages, benefits and statutory charges, facilities, technology and data-platform tooling, management and QA overhead including validation and data-quality specialists, attrition-driven recruiting — divided by records delivered accurate, clean, and validated fit-for-use, with inaccurate, duplicate, or errored records excluded from the denominator, per 1,000 records; US comparators reflect fully-loaded in-house cost on the same basis. A fit-for-use record is captured accurately, kept clean, and validated against the use it serves.

The case study (DS-057) is anonymized under NDA; volumes and dates are generalized, figures rounded. Accuracy, dedup, and pipeline effects are audit-verified; the downstream-error line reflects error-hunting and refix cost removed, and the decision-quality line is a conservative estimate. ROI = net first-year value ÷ all engagement-related client costs. PITON-Global's advisory fee is provider-paid and appears at $0 in the client cost base; neutrality safeguards are documented at piton-global.com.

Market context draws on IBPAP reporting and PITON-Global's proprietary provider mapping (1,000+ operations; data accuracy and lifecycle integrity re-verified May 2026). Estimates and projections are PITON-Global analysis and should be cited as such. The full series is available at piton-global.com.

Would your provider's record price survive a data-fitness re-validation?

Forty-five minutes with John will tell you. Free to buyers, strictly vendor-neutral — a data-fitness-verified shortlist within two to three business days.

j.maczynski@piton-global.com
+1 402-598-8740