

Data Lake / Lakehouse Health Pack

Readiness checklist — 20 checks across 5 categories. Nothing proceeds until every box is honestly ticked.

Pack	Data Lake / Lakehouse Health Pack
Document	Readiness checklist (20 items)
Use	Gate reviews · phase entry criteria · go/no-go
Date	2026-09

AnovaCloud — Engineering the AI-Native Enterprise. Frisco, Texas. This pack is a working instrument, not a brochure: complete it with your team, score honestly, and bring it to your discovery call. © 2026 AnovaCloud.

INGESTION RELIABILITY

Ingestion reliability

☐

Source register current

All sources inventoried; no pipeline reads from an unregistered source.

☐

Incremental & idempotent

Re-runs are safe; late/duplicate data handled by design.

☐

Schema-drift handling

Source schema changes detected; pipelines fail loudly, never silently.

☐

Freshness SLAs monitored

Every critical pipeline has a freshness SLA with alerting.

STORAGE HYGIENE

Storage hygiene

☐

File-size health

Small-file/compaction debt measured; compaction or auto-optimize in place.

☐

Partitioning sane

Partition columns match query patterns; skew investigated.

☐

Lifecycle & retention

Tiering and retention policies enforced; orphan data purged on schedule.

☐

Format decision recorded

Table format choice documented with rationale and migration cost noted.

TRANSFORMATION QUALITY

Transformation quality



Layer contracts enforced

Bronze↔silver↔gold contracts defined; violations blocked or alerted.



Pipeline tests in CI

Transformation logic unit-tested; integration tests on sample data.



Orchestration with SLAs

Central orchestrator; retries, alerting, and SLA attainment tracked.



Lineage to bronze

Any gold metric traceable to source within minutes.

GOVERNANCE & TRUST

Governance & trust

- ☐ **Catalog coverage**
≥90% of gold tables catalogued with owner, definition, freshness.
- ☐ **DQ gates on critical paths**
Freshness/volume/distribution checks gate the pipelines that matter.
- ☐ **PII classified & controlled**
Sensitive columns classified; access restricted and audited.
- ☐ **Change management**
Breaking changes versioned, announced, and migrated with consumers.

COST & OPS

Cost & ops

☐

Cost attribution live

Spend visible per pipeline/team; anomalies alerted.

☐

Top cost drivers owned

Each top-10 driver has an owner and an optimization target.

☐

On-call & runbooks

Rotation defined; runbooks cover the common failure modes.

☐

Serving performance

Top consumer queries meet p95 targets; caching/marts where needed.

SIGN-OFF

Checked means done, not planned

Each category owner initials when every item is complete with evidence attached. Items marked N/A need a written reason — 'not applicable' without justification scores as incomplete.

Ingestion reliability	Owner _____	Date _____
Storage hygiene	Owner _____	Date _____
Transformation quality	Owner _____	Date _____
Governance & trust	Owner _____	Date _____
Cost & ops	Owner _____	Date _____