

Databricks Engineering Pack

Discovery questionnaire — score each item honestly; low scores are the point: they show where the program will hurt.

Pack	Databricks Engineering Pack
Document	Discovery questionnaire (24 questions)
Response scales	Maturity 1–5 · Yes / Partially / No · Written
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.

HOW TO USE

Run this like a workshop, not a survey

Complete it with the stakeholders named in each section's intro. One scribe, one room (or call), 90 minutes. Score from evidence — monitoring data, documents, demos — not opinions. Any item scoring 1–2 is a program risk: it needs an owner and a close date before you commit budget. Bring the scored questionnaire to your discovery call; it compresses weeks of fact-finding into one session.

1 · WORKSPACE & UNITY CATALOG**Workspace & Unity Catalog**

Governance is the control plane.

Q1. Is Unity Catalog enabled and is it the single metastore (no legacy Hive metastore sprawl)?

1	2	3	4	5
--------------------------------	--------------------------------	--------------------------------	--------------------------------	--------------------------------

Maturity 1-5 (1=Ad hoc, 2=Repeatable, 3=Defined, 4=Managed, 5=Optimized)

Evidence: metastore inventory.

Q2. Are catalogs/schemas/volume ownership, grants, and row/column security consistently applied?

1	2	3	4	5
--------------------------------	--------------------------------	--------------------------------	--------------------------------	--------------------------------

Maturity 1-5 (1=Ad hoc, 2=Repeatable, 3=Defined, 4=Managed, 5=Optimized)

Evidence: grant audit sample.

Q3. Is cluster policy / compute governance preventing ungoverned all-purpose clusters?

☐	Yes	☐	Partially	☐	No
--------------------------	-----	--------------------------	-----------	--------------------------	----

Evidence: cluster policies + compliance rate.

Q4. Are service principals (not personal tokens) used for production jobs?

☐	Yes	☐	Partially	☐	No
--------------------------	-----	--------------------------	-----------	--------------------------	----

Evidence: credential inventory.

2 · WORKLOAD PROFILE

Workload profile

Know what you're actually running.

Q5. What is the workload mix (batch ETL, streaming, ML, ad-hoc) by DBU consumption?

Evidence: DBU breakdown by workload.

Q6. Which 10 jobs/pipelines consume the most DBUs, and are they optimized or just big?

Evidence: top-consumer list with notes.

Q7. Are interactive clusters right-sized, auto-terminating, and tagged by owner/team?

☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
----------------------------	----------------------------	----------------------------	----------------------------	----------------------------

Maturity 1-5 (1=Ad hoc, 2=Repeatable, 3=Defined, 4=Managed, 5=Optimized)

Evidence: cluster utilization + tagging compliance.

Q8. Is Photon/serverless evaluated where it fits (not assumed, not ignored)?

☐ Yes	☐ Partially	☐ No
------------------------------	------------------------------------	-----------------------------

Evidence: evaluation notes or pilot results.

3 · DELTA & PERFORMANCE

Delta & performance

Delta Lake rewards discipline and punishes neglect.

Q9. Are OPTIMIZE/Z-order (or liquid clustering) and VACUUM running on hot tables?

1	2	3	4	5
---	---	---	---	---

Maturity 1-5 (1=Ad hoc, 2=Repeatable, 3=Defined, 4=Managed, 5=Optimized)

Evidence: maintenance job history + file metrics.

Q10. Are Delta expectations / constraints enforcing data quality at write time?

☐	Yes	☐	Partially	☐	No
--------------------------	-----	--------------------------	-----------	--------------------------	----

Evidence: expectation inventory + violation handling.

Q11. Have the slowest jobs been profiled (Spark UI: skew, spill, shuffle) with fixes shipped?

1	2	3	4	5
---	---	---	---	---

Maturity 1-5 (1=Ad hoc, 2=Repeatable, 3=Defined, 4=Managed, 5=Optimized)

Evidence: profiling notes + before/after runtimes.

Q12. Is caching, broadcast, and partition strategy deliberate on the top workloads?

1	2	3	4	5
---	---	---	---	---

Maturity 1-5 (1=Ad hoc, 2=Repeatable, 3=Defined, 4=Managed, 5=Optimized)

Evidence: per-workload tuning notes.

4 · JOBS & ORCHESTRATION

Jobs & orchestration

Workflows are production software.

Q13. Are production pipelines on Workflows with retries, alerting, and SLAs — not notebooks on cron?

1	2	3	4	5
---	---	---	---	---

Maturity 1-5 (1=Ad hoc, 2=Repeatable, 3=Defined, 4=Managed, 5=Optimized)

Evidence: workflow inventory + SLA attainment.

Q14. Are jobs parameterized, version-controlled, and deployed via CI/CD (not manual notebook edits)?

1	2	3	4	5
---	---	---	---	---

Maturity 1-5 (1=Ad hoc, 2=Repeatable, 3=Defined, 4=Managed, 5=Optimized)

Evidence: repo + deployment pipeline.

Q15. Is there dependency management between jobs (no time-based coupling)?

☐	Yes	☐	Partially	☐	No
--------------------------	-----	--------------------------	-----------	--------------------------	----

Evidence: DAG review of top pipelines.

Q16. Are failed runs triaged with runbooks, and is the failed-run rate trending down?

1	2	3	4	5
---	---	---	---	---

Maturity 1-5 (1=Ad hoc, 2=Repeatable, 3=Defined, 4=Managed, 5=Optimized)

Evidence: failed-run trend + runbooks.

5 · COST & FINOPS**Cost & FinOps**

DBUs are easy to spend and hard to explain.

Q17. Is DBU spend attributed by team/job with budgets and anomaly alerts?

1	2	3	4	5
---	---	---	---	---

Maturity 1-5 (1=Ad hoc, 2=Repeatable, 3=Defined, 4=Managed, 5=Optimized)

Evidence: cost dashboards + alert rules.

Q18. Are the top-10 DBU consumers reviewed on a cadence with optimization owners?

☐	Yes	☐	Partially	☐	No
--------------------------	-----	--------------------------	-----------	--------------------------	----

Evidence: review cadence + action log.

Q19. Are committed-use / discount options evaluated against actual consumption?

☐	Yes	☐	Partially	☐	No
--------------------------	-----	--------------------------	-----------	--------------------------	----

Evidence: commitment analysis.

Q20. Is there a chargeback/showback model the business understands?

1	2	3	4	5
---	---	---	---	---

Maturity 1-5 (1=Ad hoc, 2=Repeatable, 3=Defined, 4=Managed, 5=Optimized)

Evidence: showback reports.

6 · SECURITY & COMPLIANCE

Security & compliance

The lake holds everything; guard accordingly.

Q21. Is PII classified in Unity Catalog with masking/row-filters where required?

☐

1

☐

2

☐

3

☐

4

☐

5

Maturity 1-5 (1=Ad hoc, 2=Repeatable, 3=Defined, 4=Managed, 5=Optimized)

Evidence: classification + policy sample.

Q22. Are audit logs (account + workspace) retained and reviewed?

☐

Yes

☐

Partially

☐

No

Evidence: log retention + review cadence.

Q23. Is network isolation (VPC/VNet, Private Link) configured per policy?

☐

1

☐

2

☐

3

☐

4

☐

5

Maturity 1-5 (1=Ad hoc, 2=Repeatable, 3=Defined, 4=Managed, 5=Optimized)

Evidence: network architecture.

Q24. Is there a disaster-recovery story for critical Delta tables (replication, backups)?

☐

Yes

☐

Partially

☐

No

Evidence: DR design + test record.

SCORING

Read the pattern, not just the average

Average each section. Sections averaging below 3 are your critical path — sequence the project plan so those risks close first. A single 1 on a load-bearing question (downtime budget, rollback, access control) outweighs a 4 average elsewhere. Record owners and close dates below.

Lowest-scoring section:

Single biggest risk:

Owner:

Close date:
