Duck Creek Technologies

Data Engineering Interview Questions

Ankita Gulati

Shubh Goyal

Job Details

- **Position:** Senior Data Engineer
- **Experience:** 4+ years
- **Location:** Remote
- **Work mode:** Remote
- **Compensation:** ₹25+ LPA
- **Total Rounds:** 3
- **Top Required Skills:**

1. SQL
2. PySpark / Python / Databricks
3. Cloud Data Engineering
4. ETL / Data Modeling
5. Big Data & Streaming
6. System Design

Ankita Gulati

Shubh Goyal

Round 1
Technical Discussion

1. How would you design an ingestion pipeline in AWS to capture Duck Creek policy lifecycle events (create, endorse, cancel) while ensuring event order is preserved per policy?
2. How would you model policy versions and effective periods so analytics can calculate exposure accurately?
3. How would you track versioned product-rule changes and ensure reports use the correct rule version for historical records?
4. What approach would you take to implement Change Data Capture (CDC) from a on-premise database into an AWS S3-based data lake, ensuring no missed or duplicated events?
5. Which AWS services or methods would you use to automatically detect, mask, and secure PII before storing data in analytical zones?

Ankita Gulati

Shubh Goyal

6. Duck Creek exposes SOAP-based endpoints. How would you design a connector to extract large historical datasets without overloading the source system?

7. How would you reconcile billing transactions between Duck Creek Billing and a central ledger in AWS, and how would you detect and handle mismatches?

8. How would you design pipelines and reports to ensure accurate policy exposure as of any historical date?

9. What is a data contract in data engineering, and how would you implement and enforce one between Duck Creek as the source and downstream consumers?

10. What strategy would you use to compact and organize these files in S3 for better Athena/Redshift Spectrum performance?

11. Insurers require immutable archives for audits. How would you design a tamper-proof data storage solution in AWS that supports legal hold and audit trails?

Ankita Gulati **Shubh Goyal**

Round 2
Coding & Architecure

1. How would you design an AWS architecture to process events in real-time, enrich them with reference data, and deliver both streaming features for ML models and daily aggregates for reporting?
2. What techniques would you use to guarantee exactly-once processing for Duck Creek policy events from ingestion to Redshift/Snowflake?
3. How would you design monitoring and alerting for ingestion pipelines, including key metrics and data validation checks?
4. A reporting system requires policy snapshots as-of any given date. Would you use event sourcing, SCD Type-2, or another approach? Explain your choice.
5. How would you enforce legal hold on specific company policy records stored in AWS S3 to meet regulatory requirements?
6. Company produces large XML claim payloads. How would you efficiently parse, validate, and store these for analytics while keeping data lineage intact?

Ankita Gulati

Shubh Goyal

7. How would you optimize S3 storage costs for large insurance datasets while maintaining compliance and auditability?

8. How would you securely share Company datasets across AWS accounts and business units while enforcing fine-grained access control and centralized auditing?

9. How would you detect a sudden data skew or schema change in Duck Creek data that could corrupt actuarial aggregates, and how would you mitigate it?

10. How would you implement CI/CD for Duck Creek ETL pipelines, including schema migrations, data contracts, and automated regression tests?

11. How would you automate disaster recovery (DR) to restore 30 days of Company policy data into a separate AWS account?

12. Write a function that, given endorsement records with effective_date, expiry_date, and coverage_limit, outputs non-overlapping coverage intervals in chronological order.

13. Given a policy_transactions (policy_id, txn_time, txn_type) table.
Write a query to count endorsements occurring before a cancellation for each policy.

14. How would you implement a checksum or hashing strategy to verify that a Duck Creek daily export file exactly matches the source table snapshot?

Ankita Gulati

Shubh Goyal

Round 3
HR Discussion

1. How do you balance the need for rapid delivery of Company integrations with ensuring data correctness and regulatory compliance?
2. Tell me about a time you translated a complex policy business rule into a data engineering solution. How did you resolve ambiguities?
3. How would you run an onboarding session for business users to validate the first month of Duck Creek policy data ingestion?
4. How do you stay updated with actuarial or insurance domain changes that may affect data pipelines?
5. What makes you a strong fit for working on Duck Creek integrations from both a technical and domain perspective?

Ankita Gulati

Shubh Goyal

Thank You

Best of luck with your upcoming interviews — you've got this!

Ankita Gulati

Shubh Goyal