



Data Engineering Interview Questions



Ankita Gulati

Shubh Goyal



Job Details

- **Position:** Data Engineer
- **Experience:** 3+ years
- **Location:** Bangalore
- **Work mode:** Office
- **Compensation:** ₹25+ LPA
- **Total Rounds:** 5
- **Top Required Skills:**
 1. SQL
 2. PySpark / Python
 3. Cloud Data Engineering
 4. ETL / Data Modeling
 5. Big Data & Streaming
 6. System Design

Round 1

Hiring Manager Interview

Focus: Background & Role Alignment

1. Can you introduce yourself and walk me through your career so far?
2. What are your key responsibilities in your current role?
3. How does your current work align with this role at Shopee?
 - Example: experience with Flink, Kafka, Clickhouse, and data lakes.

Round 2

Online Assessment

Coding Questions

1. Graph traversal problem: Count the number of connected components (e.g., lakes/land masses).
2. Medium-level string manipulation or sliding window problem.

Multiple-Choice Questions

3. Basics of distributed systems (e.g., CAP theorem, consistency models).
4. Real-time processing concepts (Flink, Spark).
5. Big data frameworks (Hadoop, Kafka).

Round 3

Technical Interview 1

Real-Time Data Processing

1. Walk us through a real-time data processing system you've built.

- Example: Using Flink + Kafka to process millions of events/second.

System Design

2. Design a real-time architecture to process 1 million events/second with deduplication.

- Tools: Kafka for ingestion, Flink for stream processing, Clickhouse for storage.
- Fault tolerance: Kafka replication, Flink checkpointing.

Follow-Up

3. How do you ensure low latency in high-throughput systems?

- Optimizations: Kafka partitioning, Flink operator tuning.

Round 4

Technical Interview 2

Coding Challenge

1. Write a function to deduplicate a stream of data before storing it in a database.
 - Solution: Hash map for uniqueness + discussion of Flink state management.

Database Optimization

2. How would you optimize Clickhouse for large-scale analytics?
 - Partitioning strategies, data compression, MergeTree optimizations.

Distributed Systems

3. What would you do if your Spark cluster fails to process a micro-batch on time?
 - Solutions: increase cluster resources, re-partition, optimize configs.

Round 5

Final Behavioral Round

Leadership

1. How do you align cross-functional teams during complex projects?

Conflict Resolution

2. Tell me about a time when there was a technical disagreement in your team.

- How did you mediate and resolve it?

Shopee-Specific Fit

3. Why Shopee? How does your experience align with our mission and values?

Thank You

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

