



# Data Engineering Interview Questions



Ankita Gulati

Shubh Goyal



# Job Details

- **Position:** Data Engineer II
- **Experience:** 2.5+ years
- **Location:** Bangalore
- **Work mode:** Hybrid
- **Compensation:** ₹12-14 LPA
- **Total Rounds:** 3
- **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

## Technical Screening

1. Solve an array-based problem requiring efficient traversal and manipulation techniques. Focus on time and space optimization.
2. Given a matrix, perform a Breadth-First Search (BFS) traversal. Ensure you handle edge cases such as disconnected nodes or boundary conditions.
3. Write a Python program to check if a string is a palindrome in  $O(n)$  time without extra space.
4. Implement a function in Python that finds the longest increasing subsequence in a list of integers.
5. Given a dataset of customer transactions, write a SQL query to return the top 3 customers by spending in each month.
6. Describe a time when you faced pipeline maintenance challenges. How did you resolve them?
7. How do you manage stakeholder alignment when requirements change frequently?
8. What were your biggest learnings from data quality or pipeline reliability issues?

# Round 2

## SQL & Data Engineering

1. Write a SQL query to calculate a rolling mean per product over 3 consecutive dates using window functions.
2. Given two tables, write a query using CTE, JOIN, and RANK() to fetch ranked results.
  - Follow-up: Modify the query to use a LEFT JOIN to include unmatched records.
3. Explain different types of indexes (Clustered vs. Non-clustered, Bitmap, B-Tree). When would you use each?
4. What is partitioning in databases? Differentiate between horizontal vs vertical partitioning, and explain partition pruning.
5. Compare file formats (CSV, Parquet, ORC). Which would you choose for analytics vs. raw storage, and why?
6. How would you optimize queries in BigQuery/Snowflake/Redshift?

7. What are the best practices for handling schema evolution in a data pipeline?
8. Explain how you would build a data pipeline to process real-time streaming data (e.g., with Kafka or Kinesis).
9. Python: Write a function to merge two sorted lists into a single sorted list without using built-in sort.
10. SQL: Given an employees table with id, name, salary, manager\_id, write a query to find employees earning more than their managers.
11. Expedia emphasizes “Force Simplicity”. Share a real-life example where you reduced complexity in a project and delivered clear outcomes.
12. Expedia values user-centricity. How have you built data products with user empathy and incorporated feedback?
13. What steps do you take for data cleaning before integrating with BI tools? How do you handle nulls, duplicates, and inconsistent formats?
14. How do you ensure data quality in a pipeline, and what monitoring tools have you used?

# Round 3

## HR & Behavioral

1. Tell me about yourself and why you are interested in joining Expedia.
2. Describe a time when your data pipeline or system failed in production. How did you handle it?
3. Share an example where you had to explain a complex technical concept to a non-technical stakeholder.
4. Expedia values “Force Simplicity.” Can you describe a situation where you reduced complexity and delivered a simpler, impactful solution?
5. Tell me about a conflict or disagreement with a teammate/manager and how you resolved it.
6. Where do you see yourself in the next 3–5 years, and how does this role at Expedia fit into your career goals?

*Thank You*

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

