Expedia

Data Engineering Interview Questions

Ankita Gulati

Shubh Goyal

Job Details

- **Position:** Data Engineer
- **Experience:** 5 Years
- **Location:** Bangalore, Gurgaon
- **Work mode:** Hybrid
- **Compensation:** ₹32–₹40 LPA
- **Total Rounds:** 4
- **Top Required Skills:**

1. Kafka 2.Spark 3.SQL 4.Data Pipeline Design 5. CI/CD 6.DSA (recursion, stacks) 7. Behavioral

Ankita Gulati

Shubh Goyal

Round 1

Technical

(Data PipelineDesign+Kafka Concepts)

1.Past Experience & Projects
   a.Resume walkthrough.
   b.Candidate was asked to explain end-to-end ownership of pipelines.
2.Real-Time Streaming Data Pipeline Design
   a. How would you design a streaming pipeline for processing millions of events per second?
   b.Expected Discussion: ingestion, transformation, storage, monitoring.
3.Apache Kafka Concepts (Deep Dive)
   a.Offset management: committed offsets, rebalancing, at-least-once vs exactly-once delivery.
   b.Sync vs Async commits.
   c.Partition assignment strategies (range vs round robin).
   d.Consumer groups & fault tolerance.
   e.Backpressure handling in Kafka consumers.
4.Docker for Scaling Streaming Applications
   a. How would you containerize and scale Kafka consumers with Docker?
   b.Example: Running multiple consumer containers with horizontal scaling.
5.CI/CD Deployment
   a.Discussion around automating pipeline deployments.
   b.Tools: Jenkins, GitHub Actions, ArgoCD, Kubernetes.

Round 2
Coding

1. PowerofaNumber(FastExponentiation)
   a.Writecodetocalculatex^n in minimum time complexity.
   b.Hint:Usedivide-and-conquer recursion + memoization.
   c.ExpectedTimeComplexity: O(log n).
      Example:
         Input: 2^10
         Output: 1024
2. Expression Evaluation (Infix/Postfix/Prefix)
   a.Givenanexpressionininfix/postfix/prefix, evaluate its final result.
      Input(Postfix):"231*+9 -"
      Output: -4
   b.ConceptTested:StackData Structure.
   c.Expected:Parsing+stack push/pop implementation.

Ankita Gulati

Shubh Goyal

Round 3

Technical(Spark+Kafka + SQL)

Apache Spark Questions:

1. FundamentalsofSpark execution (RDD → DAG → Stages → Tasks).
2. OOM (Out of Memory) scenarios in Spark: causes & fixes (e.g., caching, partitioning, executor memory tuning).
3. Optimizing Spark applications: use of broadcast joins, avoiding wide transformations, partitioning strategy.
4. Optimized joins in Spark (Broadcast Join vs Shuffle Join).
5. Handling Data Skewness with salting technique.

Apache Kafka Questions:

1. Fundamentals:producer,broker, consumer.
2. High throughput vs low latency trade-offs.
3. Replication factor, ISR, leader election.

SQL Questions:

1. Complex queries involving Joins + Group By.
2. Example: Find top-selling product per region, but ensure regions with no sales also appear.
3. Expected: use of LEFT JOIN + GROUP BY + COALESCE.

Ankita Gulati

Shubh Goyal

Round 4

HiringManagerDiscussion

Behavioral & Resume-Based:

1. Discussion around past experiences & projects.
2. Good & bad experiences with past employers.
3. Teamwork under tight deadlines: How do you manage delivery pressure?
4. Direct Question: “Why did you leave McKinsey & Company in just 4 months?”
    - Tested honesty + professionalism.
5. “What are you expecting in your next role?”
    - Tested alignment with Expedia’s role & career growth path.

Thank You

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

Ankita Gulati

Shubh Goyal