Nielsen Data Engineer Interview Guide – Experienced 3+

Interview Process Breakdown

Round 1: Technical

Objective: Assess the candidate's understanding of AWS services, Python programming, PySpark, and fundamental data engineering concepts.

Round 2: Techno-Managerial

Objective: Test the candidate's ability to handle complex technical scenarios, evaluate their project experience, and gauge managerial aptitude.

Round 3: HR Discussion

Objective: Discuss cultural fit, compensation, and policy-related aspects.

Detailed Insights on Each Round

Round 1: Technical

This round required a solid grasp of both theoretical concepts and practical problem-solving.

- **AWS Concepts:**
    - Infrastructure as Code (IaC): Explained as a method to manage and provision computing infrastructure using machine-readable configuration files (e.g., AWS CloudFormation or Terraform).
    - Lambda Deployment via CLI: Steps included creating a ZIP package, configuring an IAM role, deploying using AWS CLI commands, and specifying the runtime (e.g., Python, Node.js).

- **Python Programming:**
    - Difference between is (checks object identity) and == (checks value equality).
    - Techniques to remove duplicates from a list using methods like set() or list comprehensions.
    - Usage of startswith and endswith for filtering strings based on prefix or suffix.

- **PySpark and Spark Architecture:**
    - Spark Context vs. Streaming Context: Spark Context initializes core Spark functionalities, while Streaming Context manages data streams for real-time processing.
    - RDDs (Resilient Distributed Datasets): Explained as fault-tolerant collections of objects that can be operated on in parallel.
    - Difference between Map (applies a function to each element) and Reduce (aggregates data).
    - Types of joins in PySpark, including inner, outer, left, and right joins.

- DAG Scheduler: Discussed its role in breaking jobs into stages and tasks, optimizing execution flow.

- **Programming Task:**

   Wrote a Python program to count character frequencies in a string while treating uppercase and lowercase letters as distinct.

Round 2: Techno-Managerial

Led by the Director of Engineering, this round was intensive and focused on AWS, EMR, and project management.

- **EMR Expertise:**

  - Bootstrap Actions: Custom scripts run during the startup of an EMR cluster.

  - Steps to create an EMR cluster: Configuring instance types, roles, and security groups; adding steps; and setting termination conditions.

  - Making EMR clusters resilient: Discussed strategies like using EMR fleets for dynamic resource allocation and auto-scaling for fault tolerance.

- **Scenario-Based AWS Questions:**

  - Designed a solution involving multiple services like EC2, S3, Glue, EMR, Lambda, Redshift, and Athena to handle end-to-end data workflows.

  - Example: Restricting access to an S3 bucket for a specific EC2 instance using bucket policies and instance profiles.

- **AWS KMS (Key Management Service):**

   Purpose: Encrypting data using managed keys for secure storage and processing.

- **Portfolio Review:**

  - LinkedIn and GitHub profiles were assessed for certifications, project quality, and activity.

  - Discussed implementation details of portfolio projects and answered in-depth cross-questions about architecture and decisions.

Round 3: HR Discussion

This round revolved around organizational fit and finalizing compensation.

- **Team Culture Discussion:**

   Explored team dynamics, collaborative approaches, and alignment with the company's vision.

- **Policy Discussions:**

   Clarified leave entitlements, holiday policies, and variable pay structure.

- **Salary Negotiation:**

  Negotiated base salary, variable components, and other perks based on role expectations and market standards.

4. Example Questions for Each Round

Round 1 (Technical):

1. Explain how Infrastructure as Code (IaC) works in AWS and its advantages.
2. Write a Python program to remove duplicate elements from a list while preserving the original order.
3. Describe the role of a DAG Scheduler in PySpark.
4. What are the key differences between Map and Reduce in Spark?

Round 2 (Techno-Managerial):

1. How do you ensure fault tolerance when processing large datasets in EMR?
2. What is the role of AWS KMS in securing sensitive data?
3. Explain how you would configure an S3 bucket policy to allow access only from a specific EC2 instance.
4. Describe a challenging project and how you resolved a critical technical issue.

Round 3 (HR Discussion):

1. What do you value most in team collaboration and culture?
2. Are you comfortable with the variable pay structure, and what are your expectations for the base salary?

Glassdoor Nielsen Review –

https://www.glassdoor.co.in/Reviews/Nielsen-Reviews-E3776.htm

Nielsen Careers –

https://www.nielsen.com/careers/careers/

Subscribe to my YouTube Channel for Free Data Engineering Content –

https://www.youtube.com/@shubhamwadekar27

Connect with me here –

https://bento.me/shubhamwadekar

Checkout more Interview Preparation Material on –

https://topmate.io/shubham_wadekar