

Assessment Test: Full Stack AI Developer

Candidate: Shekhar_nonlatex

Introduction

Welcome to this comprehensive technical assessment designed specifically for the Full Stack AI Developer role at Akro. Your skills and expertise will be evaluated through a series of questions covering technical skills, problem-solving, and real-world scenarios.

Instructions

Please read each question carefully and respond accordingly. Multiple-choice questions require you to select one correct answer from the provided options. Open-ended questions allow for your thoughtful and detailed responses. You can use any necessary tools or resources during this assessment.

Technical Skills

Q1. What is the primary programming language used in Apache Airflow?

- a) Python
- b) Scala
- c) Java
- d) SQL

Q2. Describe a scenario where you would use Apache Camel to integrate different data sources. Please provide a concise and clear explanation.

Data Engineering and Pipelines

Q1. What is the primary purpose of using Cassandra in big data applications?

- a) Distributed database for real-time analytics
- b) Scalable NoSQL database for large-scale data storage

- c) Cloud-based object store for data processing
- d) Real-time event processing and stream processing

Q2. Design a simple ETL (Extract, Transform, Load) pipeline using PySpark to process data from a CSV file. Please provide a step-by-step explanation of your approach.

AI and Machine Learning

Q1. What is the primary benefit of using Hadoop in AI and machine learning applications?

- a) Scalability for large-scale data processing
- b) Ease of use for data scientists
- c) Flexibility for integrating with various ML frameworks
- d) Cost-effectiveness for storing large datasets

Q2. Describe a scenario where you would use SQL to optimize the performance of an AI-powered recommendation system. Please provide a concise and clear explanation, highlighting any relevant SQL queries or algorithms.