Project Title: Real-time E-commerce Data Pipeline and Analytics Dashboard

Project Overview: Students will build a comprehensive e-commerce data pipeline that collects, processes, and visualizes real-time data from a simulated online store. The system will use various AWS services, incorporate data processing with Python, store data in a PostgreSQL database, and visualize insights using Tableau and Power BI.

Key Components:

1. Data Generation and Ingestion:
   * Create a Python script to simulate real-time e-commerce data (orders, user interactions, inventory updates).
   * Use AWS Kinesis for real-time data streaming.
2. Data Processing and Storage:
   * Implement AWS Lambda functions (using Python) to process incoming data streams.
   * Store processed data in Amazon RDS (PostgreSQL).
3. Data Structures and Algorithms:
   * Implement a custom data structure (e.g., a modified trie) for efficient product categorization and search.
   * Develop algorithms for real-time aggregations and basic trend detection in sales data.
4. API Development:
   * Create a RESTful API using AWS API Gateway and Lambda to expose data and insights.
5. Visualization and Reporting:
   * Develop interactive dashboards using Tableau and Power BI, connected to the PostgreSQL database.
   * Create real-time monitoring dashboards using AWS CloudWatch.

Project Tasks:

1. Setup and Configuration:
   * Set up AWS environment (IAM, VPC, EC2, RDS, Kinesis, Lambda, S3).
   * Install and configure PostgreSQL on Amazon RDS.
2. Data Generation and Ingestion:
   * Develop a Python script to generate realistic e-commerce data.
   * Configure AWS Kinesis Data Streams for data ingestion.
3. Data Processing:
   * Implement AWS Lambda functions for data transformation and enrichment.
   * Use Python to clean and preprocess data before storage.
4. Database Design and Implementation:
   * Design a normalized database schema for the e-commerce platform.
   * Implement the schema in PostgreSQL, including appropriate indexes and constraints.
5. Advanced Data Structures and Algorithms:
   * Implement a custom trie-based data structure for product categorization and search.
   * Develop algorithms for real-time aggregations (e.g., moving averages, top N products) using efficient data structures like heaps or balanced binary search trees.
6. API Development:
   * Design and implement a RESTful API using AWS API Gateway and Lambda.
   * Implement authentication and rate limiting for the API.
7. Visualization and Reporting:
   * Create interactive dashboards in Tableau and Power BI, focusing on:
     + Real-time sales performance
     + Inventory levels and turnover rates
     + Customer behavior patterns
     + Geographical sales distribution
   * Develop real-time monitoring solutions using AWS CloudWatch for system performance and data pipeline health.
8. Performance Optimization:
   * Optimize database queries using SQL techniques (indexing, partitioning).
   * Implement caching mechanisms using AWS ElastiCache to improve API response times.
9. Data Pipeline Monitoring and Error Handling:
   * Implement comprehensive error handling and logging throughout the pipeline.
   * Create alerts and notifications for critical errors or data anomalies.
10. Documentation and Presentation:
    * Document the entire system architecture and data flow.
    * Prepare a presentation showcasing the platform's capabilities and potential business impact.