**Use case: Virtual Server Monitoring and Performance Optimization**

**Problem Statement**: XYZ Corporation, a global enterprise, operates a distributed cloud infrastructure comprising numerous virtual servers that host business-critical applications. However, the organization faces challenges in effectively monitoring server performance, detecting anomalies, and proactively managing resource utilization. Real-time insights into server health and availability are limited due to the lack of a robust data engineering pipeline.

To improve operational efficiency and prevent downtime, XYZ Corporation aims to build an end-to-end data pipeline that ingests, processes, and visualizes server performance data. This pipeline will enable proactive issue resolution, optimized resource utilization, and enhanced server performance management.

**Data Description:**  
Teams are provided with an Excel file containing the following sheets:

1. **Server\_Metadata**: Contains 100 records detailing server configurations, cluster assignments, and administrator contact details.
2. **Server\_Performance\_Station1**: Contains 3,000 records of performance logs captured from Station 1.

It includes additional fields such as Config\_Version, Last\_Patch\_Date, and Deployment\_Token which are not relevant for visualization.

1. **Server\_Performance\_Station2**: Contains 2,000 records of performance logs captured from Station 2.

Both performance datasets include intentional null values.

Here are the key components of the problem statement:

1. **Data Ingestion**:
   * **Preferred**: Ingest data into Azure Cloud using services like Azure Data Factory or Azure Synapse.
   * **Alternative**: If Azure access is not available, implement the data ingestion locally using Python notebooks or similar tools.
   * Ensure data quality and handle any data anomalies during ingestion.
2. **Data Transformation**:
   * Transform the raw server logs to create meaningful insights.
   * Calculate CPU Utilization, Memory Usage, and other relevant metrics.
   * Enrich the dataset with metadata such as **server location, OS type, and instance size**.
   * Store the transformed data into a structured data store.
3. **Power BI Dashboard**:
   * Create a Power BI dashboard using the structured data store to visualize server performance trends, uptime statistics, and resource utilization.
   * Include interactive visuals, filters, and KPIs.

**Guidelines for Submission**

**The use of Azure Cloud is recommended but not mandatory. You may implement the full pipeline locally if Azure trial or resources are unavailable. (Preferred would be using Azure Cloud services)**

* A working data pipeline (code/scripts or notebook) that performs ingestion, cleaning, transformation, and loading.

(Notebook to be clean and have proper markdowns)

* A joined and cleaned dataset ready for visualization.
* A Power BI report/dashboard (.pbix or screenshot) with meaningful KPIs and interactive visuals.
* Detailed documentation(.docx) or notebook explaining:
  + Architecture diagram, Data Flow Diagram and Data Model
  + Assumptions made
  + Data cleaning strategy
  + Transformation logic
  + Visual insights

**Evaluation Criteria**

Your work will be evaluated based on the following criteria:

1. **Functionality**:
   * Does the solution meet the problem statement requirements?
   * Is the data pipeline functioning correctly?
2. **Scalability and Performance**:
   * Consider the scalability of your chosen storage solution.
   * Evaluate the efficiency of data transformation processes.
3. **Creativity and Attention to Detail**:
   * Did you explore additional features beyond the basic requirements?
   * Is the Power BI dashboard providing proper insights, visually appealing and user-friendly?

Remember, this project is an opportunity to showcase your skills, creativity, and problem-solving abilities. Candidates are expected to showcase their analytical, logical thinking, and deduction skills throughout this challenge. Your approach to the problem, as well as your ability to present and defend your findings, will be central to the evaluation process.

Please ensure the work is your original work. Any kind of plagiarism or taking help from others will have adverse consequences. Refrain from using ChatGPT, Gemini or any other AI tool, this will lead to immediate disqualification. We strictly do not tolerate any kind of malpractice or mis behaviour.

**Best of Luck!**

**Team NeoStats.**