Z2104 - End Semeseter Examination - Part 2

Course: Data Engineering and Data Mangement - Z2104
Duration: 50 minutes
Total Marks: 50
Each of you may have subtle variation in your datasets.
Do not attempt to copy another student's solution !!

July 2, 2025

1 Examination Tasks

1.1 Task 1: Environment Setup (10 marks)

Objective: Deploy the Docker environment and verify all services are running.

1. Download the provided `docker-compose.yml` file
2. Deploy the Docker environment using Docker Compose
3. Verify that all four services (MySQL, Grafana, InfluxDB, Node-RED) are running
4. Access the Grafana web interface and take a screenshot

Deliverables:

- Screenshot of `docker ps` command output showing running containers
- Screenshot of Grafana login page (http://localhost:3000)

1.2 Task 2: Database Access and Exploration (15 marks)

Objective: Connect to the MySQL database and explore the dataset structure.

1. Connect to the MySQL database using the provided credentials
2. List all tables in the 'workshop' database
3. Execute the provided sample queries to understand the data structure
4. Write a query to find the top 5 products by total sales amount

Sample Query to Execute:

```
SELECT category, COUNT(*) as product_count,
       AVG(total_amount) as avg_sales
FROM sales_data
WHERE sale_date >= '2024-01-01'
GROUP BY category
ORDER BY avg_sales DESC;
```

Deliverables:

- Screenshot of database connection
- Output of the sample query
- Your custom query results for top 5 products

1.3 Task 3: Data Curation (10 marks)

Objective: Prepare and analyze data for visualization.

1. Create a query that shows monthly sales trends for the current year
2. Calculate the total revenue by region
3. Identify the employee performance summary (average performance score by department)

Required Queries:

1. Monthly sales aggregation
2. Regional revenue analysis
3. Employee performance by department

Deliverables:

- SQL queries used for data curation
- Screenshots of query results

1.4 Task 4: Visualization Creation (15 marks)

Objective: Create three different types of visualizations in Grafana.

1. Configure MySQL as a data source in Grafana
2. Create a **Bar Chart** showing sales by product category
3. Create a **Time Series** chart showing sales trends over time
4. Create a **Stat Panel** showing total revenue
5. Apply appropriate titles, labels, and formatting to all charts

Visualization Requirements:

- Bar Chart: Sales by Category (use sales_data table)
- Time Series: Monthly sales trend (use sale_date field)
- Stat Panel: Total revenue with currency formatting
- All charts must have descriptive titles and proper axis labels

Deliverables:

- Screenshot of Grafana data source configuration
- Screenshots of all three completed visualizations
- Export of the dashboard JSON (optional bonus)

2 Marking Scheme

Task	Marks	Criteria
Environment Setup	10	Docker deployment (4 marks) Service verification (3 marks) Grafana access (3 marks)
Database Operations	15	Database connection (3 marks) Table exploration (4 marks) Sample query execution (4 marks) Custom query (4 marks)
Data Curation	10	Monthly sales query (4 marks) Regional analysis (3 marks) Employee performance query (3 marks)
Visualization	15	Data source setup (3 marks) Bar chart (4 marks) Time series chart (4 marks) Stat panel (2 marks) Formatting & labels (2 marks)
Total	**50**	

3 Submission Requirements

1. Create a single PDF document containing all screenshots and query results
2. Include your SQL queries in a separate text file
3. Submit both files before the end of exam through a github repo which is shared with Sathish-IITMZ
4. Name your files: `StudentID_DataViz_Assignment.pdf` and `StudentID_Queries.txt`

4 Resources Provided

- `docker-compose.yml` - Docker environment configuration
- `assignment_dataset.sql` - Pre-loaded sample data
- `assignment_guide.md` - Step-by-step instructions

5 Important Notes

- **Each of you may have subtle variation in your datasets. Do not attempt to copy another student's solution !!**
- Not all code given in this assignment may be completely correct. If there are errors in the code execution use your judgment to debug the code.
- Ensure Docker Desktop is running before starting
- Use only the tools and techniques covered in class
- Focus on clarity and accuracy in your visualizations
- Partial marks will be awarded for incomplete but correct attempts. Make sure you take a screehshot and submit it in github for partial credits