BrightLight Data Analytics

Research Assignment 1

Instructions:
Research and answer the following questions.
All answers must be **handwritten**, scanned, and emailed to:
■ **rofhiwa@brightlighttutorials.co.za**

Due Date: Monday, 20 October 2025 — 23:59
✍■ *Answers must be handwritten.*

Section A: Database Fundamentals

1. What are the main types of databases?
2. What is a Relational Database Management System (RDBMS)?
3. What is a primary key and a foreign key in a database?
4. What is database normalization and why is it important?
5. What is a database schema?
6. Differentiate between structured, semi-structured, and unstructured data.
7. What is the difference between a Fact Table and a Dimension Table in a data warehouse?
8. What is a data model, and why is it important in database design?
9. Explain the difference between a database, a data warehouse, and a data lake.
10. What is a data mart, and how does it differ from a data warehouse?

Section B: SQL and Data Processing

11. What is a query language, and why is SQL the most commonly used?
12. What are indexes in databases, and how do they improve performance?
13. What are transactions in databases, and what are the ACID properties?
14. What is a database engine, and how does it impact performance?
15. What are views, stored procedures, and triggers in SQL?
16. Differentiate between ETL (Extract, Transform, Load) and ELT (Extract, Load, Transform).
17. Differentiate between batch processing and stream processing in data pipelines.
18. Explain what a join is in SQL and list different types of joins with examples.
19. What is referential integrity, and why is it important in relational databases?
20. How does data redundancy affect database performance and storage?

Section C: Data Management and Analytics Concepts

21. How does cloud database management differ from on-premise databases?
22. What is data governance, and why is it important in data management?

23. What is data integrity, and how can it be maintained?

24. What is data quality, and why is it critical for analytics?

25. Explain the role of a Data Analyst in managing and analyzing database information.

26. What are the key responsibilities of a Database Administrator (DBA)?

27. What are the main steps involved in designing a data pipeline?

28. What are some common challenges in managing large-scale databases?

29. What are some popular database platforms (e.g., MySQL, Snowflake, PostgreSQL, Oracle) and their use cases?

30. What are the main data storage formats used in analytics (e.g., CSV, Parquet, JSON, Avro)?