**COMP 263 – Mid Term: Answer Key**

|  |  |  |
| --- | --- | --- |
| # | Question Summary | Correct Answer |
| 1 | SQL vs RDBMS | B) SQL is a query language; RDBMS is the software that runs and manages the data |
| 2 | SQL DDL keyword | C) CREATE |
| 3 | SQL keyword to retrieve data | B) SELECT |
| 4 | ACID – Consistency meaning | B) The database remains valid before and after a transaction |
| 5 | CAP Theorem components | C) Consistency, Availability, Partition Tolerance |
| 6 | SQLite unique feature | C) It stores all data in a single local file |
| 7 | NoSQL reason (Strozzi) | B) SQL requires complex licenses and lacks flexibility |
| 8 | Purpose of metadata | B) It tracks data ownership and freshness |
| 9 | Not a data quality dimension | C) Speed |
| 10 | Unit mismatch (kg vs lbs) issue | B) Validity |
| 11 | 2000°C reading issue | B) Accuracy |
| 12 | Reason for UTC timestamps | B) To ensure all records use a consistent time zone |
| 13 | Importance of filtering | A) Removes irrelevant or incorrect records |
| 14 | Purpose of map() | B) Transform each element of an array and return a new one |
| 15 | Purpose of filter() | B) Select only elements meeting certain conditions |
| 16 | Purpose of reduce() | A) Combines all elements into one summarized result |
| 17 | Goal of ETL pipeline | B) Extract, clean, and load data into a warehouse |
| 18 | When ELT transformation occurs | B) After data is loaded into the target system |
| 19 | Descriptive – Metadata improves data quality | *(Answers will vary, but should mention that metadata such as author, last\_sync, and timestamp improves consistency, traceability, and data accuracy during synchronization.)* |
| 20 | Draw a simple data pipeline | *(Diagram should include: Data Sources → Ingestion → Processing → Storage → Visualization, with short labels or notes on each stage.)* |