Quiz 9

Databases

18 May 2026

Full name: .....

You have 10 minutes to answer this quiz, directly on this sheet of paper. No electronic devices or material of any kind is allowed. Do not forget to add your name above. Every question amounts to 2 points; the quiz is graded out of 10 points.

Questions

Q1. Circle the only correct answer.

Given a pandas `DataFrame` `reservation`, which expression corresponds to the relational-algebra operation $\sigma_{\texttt{nights}>3}(\texttt{Reservation})$?

a. `reservation[['nights']] > 3`

b. `reservation.where('nights > 3')`

c. `reservation[reservation.nights > 3]`

d. `reservation.groupby('nights').filter(> 3)`

Q2. Circle all correct answers (there may be more than one).

For which workloads is an SQL database management system (DBMS) generally a **better** choice than pandas?

a. Many concurrent users updating the same data, with transactional guarantees.

b. In-memory feature engineering before training a scikit-learn model.

c. Indexed lookups in a table of one billion rows that does not fit in RAM.

d. Enforcing integrity constraints such as `FOREIGN KEY` or `NOT NULL`.

Q3. Consider the following pandas expression, intended to keep the reservations of guest 2 that last more than three nights:

```
reservation[reservation.guest == 2 and reservation.nights > 3]
```

Explain briefly why this code is **wrong**, and write a correct version.

.....

.....

.....

Q4. Circle the only correct answer.

What does the following pandas snippet return?

```
reservation.groupby('room')[['nights']].sum()
```

a. the number of reservations per room

b. the total number of nights booked, per room

c. the average number of nights per room

d. the list of rooms with more than one reservation

Q5. Circle the only correct answer.

In a typical AI pipeline that uses both an SQL DBMS and pandas, what is the main benefit of writing the `JOIN`s, `WHERE` filters and `GROUP BY` aggregations inside the SQL string passed to `pd.read_sql_query` rather than loading the whole tables into pandas first?

a. Python cannot perform joins or aggregations on `DataFrame`s.

b. The DBMS uses indexes and a query optimizer, and only the (much smaller) result is transferred into Python memory.

c. Pandas can only read CSV files, not SQL databases.

d. It avoids the need for a Python interpreter on the server.