Subject number : XXXXXXXXX

Sujet corrigé - 5 - Data Storage and NoSQL

This is the **question paper**. It is **NOT** the answer sheet.
Please check that the number on your question paper matches the number on your answer sheet. To complete the answer sheet correctly, you must:

- use a **black** ink pen
- shade in the boxes **completely without going over the edges**
- if you make a mistake, erase the box with a whiteout ("Tipp-Ex"), **but do not redraw it**
- every question has a **unique correct answer**.

Box correctly ticked

1 A (B) C D E

Box incorrectly ticked

1 A (B) C D E

1 **Which of the following is NOT a storage model?**
1 Point - Only one correct choice

- ☐ **A.** Flat files
- ☐ **B.** Columnar
- ☐ **C.** Hierarchical
- ☒ **D.** Relational

> Relational (algebra) is a formal framework for querying relational databases, not a storage model.

2 **What is the main advantage of in-memory storage over on-disk storage?**
1 Point - Only one correct choice

- ☐ **A.** More predictable long-term durability
- ☒ **B.** Faster read/write operations
- ☐ **C.** Better support for large-capacity archives
- ☐ **D.** Lower hardware cost for large datasets

> In-memory storage (RAM) is significantly faster than on-disk storage, but it is volatile and more expensive.

3 **Why is Parquet preferred for big data analytics?**
1 Point - Only one correct choice

- ☐ **A.** It is human-readable and easy to inspect
- ☒ **B.** It is columnar and highly compressed
- ☐ **C.** It is optimized for OLTP workloads
- ☐ **D.** It eliminates the need for schema management

> Parquet is optimized for analytical queries due to its columnar format and efficient compression.

4 **What is the primary limitation of CSV files?**
1 Point - Only one correct choice

- ☐ **A.** They are human-readable but inefficient for analytics
- ☒ **B.** They treat everything as text and are easy to corrupt
- ☐ **C.** They enforce strict typing rules across all rows
- ☐ **D.** They require specialized software for basic use

> CSV files treat all data as strings, lack schema enforcement, and are prone to errors if not handled carefully.

5 **Which of the following is a characteristic of NoSQL databases?**
1 Point - Only one correct choice

- ☐ **A.** Strict schema requirements
- ☒ **B.** Horizontal scalability
- ☐ **C.** ACID guarantees for all operations
- ☐ **D.** Limitation to analytical workloads

> NoSQL databases are designed for horizontal scalability and flexible schema, unlike traditional SQL databases, They can be used for both OLAP and OTP systems.

6 **What is a key difference between OLTP and OLAP systems?**
1 Point - Only one correct choice

- [ ] **A.** OLTP relies on columnar formats; OLAP relies on row-oriented formats
- [x] **B.** OLTP handles small, frequent transactions; OLAP handles complex queries on large datasets
- [ ] **C.** OLTP uses NoSQL, OLAP uses SQL
- [ ] **D.** OLTP systems are always slower due to higher consistency guarantees

> OLTP (Online Transaction Processing) focuses on fast, simple transactions, while OLAP (Online Analytical Processing) is optimized for complex, large-scale queries.

7 **What is the primary purpose of the "aggregate" concept in NoSQL databases?**
1 Point - Only one correct choice

- [ ] **A.** To enforce uniform structure across all collections
- [ ] **B.** To provide a more rigid alternative to SQL tables
- [x] **C.** To serve as the atomic unit for distribution, replication, and operations
- [ ] **D.** To ensure data is physically stored in a single location

> Aggregates are the basic unit for distribution, replication, and operations in NoSQL, allowing for flexible and scalable data management.

8 **In MongoDB, what is the equivalent of a "row" in a relational database?**
1 Point - Only one correct choice

- [ ] **A.** A collection
- [x] **B.** A document
- [ ] **C.** A field
- [ ] **D.** A database

> In MongoDB, a document is analogous to a row in a relational database, representing a single record.

9 **What does the CAP theorem state about distributed databases?**
1 Point - Only one correct choice

- [ ] **A.** They can achieve Consistency, Availability, and Partition tolerance simultaneously
- [x] **B.** They can only guarantee two out of Consistency, Availability, and Partition tolerance
- [ ] **C.** They must prioritize Consistency over Availability
- [ ] **D.** They should always avoid Partition tolerance

> The CAP theorem states that distributed databases can only guarantee two of the three properties: Consistency, Availability, and Partition tolerance; trade-offs must be made.

10 **Why might a developer choose polyglot persistence in a data storage strategy?**
1 Point - Only one correct choice

- [ ] **A.** To simplify system maintenance by standardizing data management
- [x] **B.** To match different data and workload requirements with the most suitable storage technology
- [ ] **C.** To ensure consistent data modeling rules across all components of the system
- [ ] **D.** To reduce operational overhead by limiting the number of storage technologies in use

> Polyglot persistence selects the best storage technology for each data type or workload, rather than forcing a single model for everything.