1 OLTP vs OLAP

For each workloads, choose whether it's better characterized as Online Transaction Processing (OLTP) or Online Analytical Processing (OLAP).

(a) A social media site with millions of users needs to track all the "likes" and "dislikes" that each post receives.

(b) An online book store needs to aggregate and analyze its users book purchases by genre over the last eight months.

(c) An multiplayer online game has added updated areas to its map and now wants to assess how users behave in those areas, and how user playtime has changed as a result.

2 Scaling

(a) A small startup realizes that its current database can't sustain their growing workloads. Given that these workloads involve lot of write but few reads, should it invest in more partitioning or more replication?

(b) A mechanical failure causes some of the startup's database machines to permanently crash, losing data in the process. If the startup wants to prevent similar losses in the future, should it invest more in partitioning or more replication?

3 BASE

(a) Database designer Doug is annoyed with his distributed database because for some time after issuing a write, all his reads return different values. Does this violate any of the BASE properties?

Which properties of BASE do these scenarios violate?

(b) All reads and writes always have the same views of data, but they sometimes respond to valid inputs with timeout errors.

(c) Writes only propagate to 3 replicas, but the system has 5 replicas of each piece of data.

(d) An empty database that has never been populated responds to a read query on some specific key with the message "Error: key nonexistent!"

4 Key-Value Stores

Database Doug now has the following tables:
  Sales (sid, date, quantity, customer, product)
  Product (pid, name, price)
  Customer (cid, name, address)
Sales data is stored with Key = sid, Value = entire Sales record, partitioned on hash function h and replicated across 3 servers.

(a) Describe how the operation get(sid1) would be executed. (Assume a Sale with sid1 exists in the data).

(b) Describe how the operation put(sid2, saleRecord) would be executed.

(c) After put(sid2, saleRecord) is executed, is it guaranteed that every app will be able to access that new Sale data?

5 JSON

1. Convert the following relational table into a JSON document.

Players

name	debut	goals
Tony	10/12/09	43
Katy	1/20/14	22

2. Convert the following JSON document into two relational tables, Players(name, debut) and Goals(name, goals).

{"players": [
  {"name": "Abby", "debut": "10/12/09",
    "goals": 43},
  {"name": "Babby", "debut": "1/20/14",
    "goals": 22},
  {"name": "Cabby", "debut": "1/21/14",
    "goals": 23}
  ]
}

6 Mongo Query Language (MQL)

For the entire question, consider the MongoDB collection **teams** with the following fields:

- teamId (int)

- divisionId (int)

- stadiumCapacity (int)

- wins (int)

- losses (int)

- coach (string)

- captain (string)

1. Using MQL, write a query to fetch the following: Find the **coach** and **captain** of all teams from **division** 1 with at least 10 **wins**, sorted by **coach** DESC and ties broken by **captain** ASC.

2. Translate the following SQL query into an MQL query:

```
SELECT divisionId AS div, MAX(wins) AS maxWins
FROM team WHERE stadiumCapacity >= 20000
GROUP BY divisionId
SORT BY MAX(wins), COUNT(*) DESC;
```