1. A banking system needs to ensure no transaction loss or inconsistency even in case of power failure. Which property of DBMS is most relevant?
 A. Concurrency
 B. Durability
 C. Non-redundancy
 D. Efficiency

Answer: B — Durability ensures transaction results persist despite failures.

2. A distributed database spanning multiple continents suffers from network partitioning. Which CAP properties can realistically be preserved simultaneously?
 A. Consistency and Availability
 B. Availability and Partition Tolerance
 C. Consistency and Partition Tolerance
 D. All three

Answer: B and C — CAP theorem forbids all three; systems must trade off between C-A-P.

3. If a company must scale rapidly with heterogeneous data formats (JSON, CSV, sensor data), which database type suits best?
 A. Relational DBMS
 B. Graph Database
 C. Document Store
 D. Column Store

Answer: C — Document DBs handle heterogeneous, schema-flexible data effectively.

4. In an e-commerce platform, which anomalies can occur if normalization is not properly applied?
 A. Redundant data
 B. Data inconsistency
 C. Faster query performance
 D. Deletion anomaly

Answer: A, B, D — Denormalization causes redundancy, inconsistency, and anomalies.

5. A transaction scheduler enforces 2-phase locking. What is the main trade-off?
 A. Higher throughput, weaker isolation
 B. Lower concurrency, stronger isolation
 C. No deadlocks
 D. Better recovery control

Answer: B — 2PL enforces serializability but limits concurrency.

6. Which modeling method would best represent a social network with relationships and multiple entity types?
 A. E-R Model
 B. UML Class Diagram
 C. Graph Model
 D. Relational Schema

Answer: C — Graph models handle rich relationship semantics.

7. You are designing a high-read, low-write analytical system. Which data organization strategies help performance?
 A. Column-oriented storage
 B. Row-oriented storage
 C. Data compression (e.g., run-length encoding)
 D. Schema normalization

Answer: A, C — Columnar storage with compression optimizes read-heavy workloads.

8. A company needs real-time analytics from distributed logs. Which system type fits best?
 A. Document Database
 B. Graph Database
 C. Column Store
 D. Distributed Key-Value Store

Answer: D — Key-value stores scale horizontally for fast distributed reads/writes.

9. Which relational operation would you use to retrieve all employees working in both HR and Finance departments?
 A. Projection
 B. Union
 C. Intersection
 D. Join

Answer: C — Intersection retrieves common tuples across relations.

10. A database uses timestamp ordering for concurrency control. What scenario can still cause rollback?
 A. Conflicting writes with older timestamps
 B. Serial execution

C. Read-only transactions

D. Use of strict two-phase locking

Answer: A — Timestamp conflicts may cause transaction aborts/rollbacks.

11. You must migrate from a relational schema to a graph database. What adjustments are likely needed?

A. Introduce node/edge identifiers

B. Flatten recursive relationships

C. Move from fixed schema to property-based model

D. Use tuple-based joins instead of traversal

Answer: A, C — Graph DBs rely on nodes/edges with flexible properties.

12. In a distributed database, replication transparency ensures:

A. Each copy must be identical

B. User queries behave the same regardless of replica location

C. Updates propagate only manually

D. Network failures are hidden

Answer: B — Replication transparency hides replica management from users.

13. When is a weakly consistent system acceptable?

A. In financial ledgers

B. In live stock trading

C. In social media feeds

D. In bank transfer logs

Answer: C — Eventual consistency suffices for social applications.

14. A sensor data platform stores billions of readings with variable schema. Which features are beneficial?

A. Schema independence

B. Versioned data

C. Strict normalization

D. Scalability

Answer: A, D — Schema flexibility and scalability matter most.

15. Which join will preserve all tuples from the left relation even if no match exists?
 A. Inner Join
 B. Natural Join
 C. Left Outer Join
 D. Semi Join

Answer: C — Left outer join includes unmatched left tuples.

16. You observe semantic overloading in a relational schema. What is the probable issue?
 A. Too many tables
 B. One attribute used to store multiple meanings
 C. High redundancy
 D. Weak relationships

Answer: B — Semantic overloading means attributes lose precise meaning.

17. For an XML-based data interchange between systems, which standards are most critical?
 A. XQuery
 B. XML Schema
 C. XSLT
 D. XPath

Answer: A, B, C — Schema defines structure; XQuery/XSLT handle querying and transformation.

18. In a MapReduce job, which phase is responsible for combining values with identical keys?
 A. Map
 B. Shuffle
 C. Reduce
 D. Split

Answer: C — Reduce phase aggregates intermediate key-value pairs.

19. A graph DB model supports weighted edges and multiple edge types. Which model best fits?
 A. Property Graph
 B. Simple Directed Graph
 C. Hypergraph
 D. Nested Graph

Answer: A, D — Property graphs support metadata; nested graphs handle hierarchy.

20. Which factor primarily drives the need for distributed databases?
 A. Query simplicity
 B. Redundancy reduction
 C. Scalability and fault tolerance
 D. Security

Answer: C — Distributed DBs are designed for scalable and fault-tolerant architectures.

21. A hospital database frequently updates patient records but rarely deletes them. Which storage strategy will minimize write amplification and improve access?
 A. Row-oriented storage
 B. Column-oriented storage
 C. Graph storage
 D. Document storage

Answer: A — Row stores handle frequent updates efficiently since tuples are co-located.

22. A relational schema is normalized into 3NF. However, query performance drops. What are valid reasons?
 A. Excessive joins due to decomposed tables
 B. Larger indexes
 C. Reduced data redundancy
 D. Poor caching locality

Answer: A, D — More joins and fragmented data reduce performance despite theoretical soundness.

23. In a distributed system using eventual consistency, which statement is TRUE?
 A. All nodes always see the same data instantly
 B. Some reads may return stale data temporarily
 C. Writes are serialized globally
 D. Data replication is synchronous

Answer: B — Eventual consistency allows temporary divergence.

24. If a database designer mistakenly stores "TotalSalary" as an attribute derived from other columns, what anomaly may arise?
 A. Update anomaly
 B. Insertion anomaly

C. Deletion anomaly

D. Redundancy anomaly

Answer: A — Derived fields must be recomputed; inconsistent updates cause anomalies.

25. A graph traversal query performs poorly. What optimizations could help?

A. Use adjacency lists instead of matrices

B. Store frequently accessed edges in cache

C. Introduce redundant edge labels

D. Denormalize graph structure

Answer: A, B — Adjacency lists and caching reduce traversal cost.

26. Which feature of DBMS ensures rollback to last consistent state after a crash?

A. Scheduler

B. Recovery Manager

C. Buffer Manager

D. Query Parser

Answer: B — Recovery Manager handles log-based rollback and restart.

27. A system requires high concurrency but is tolerant to temporary inconsistencies. Which isolation level fits?

A. Serializable

B. Repeatable Read

C. Read Committed

D. Read Uncommitted

Answer: C — Balances concurrency with moderate isolation.

28. A column store compresses data using multiple encodings. Which properties influence encoding choice?

A. Value distribution

B. Column cardinality

C. Query workload

D. Tuple size

Answer: A, B, C — Encoding must suit data patterns and access type.

29. Which XML concept is loosely analogous to SQL's SELECT-FROM-WHERE clause?

A. XSLT

B. XPath

C. FLWOR Expression

D. XML Schema

Answer: C — FLWOR defines query expressions like SQL.

30. In a NoSQL key-value store, you observe skewed load on some nodes. What techniques can mitigate this?
 A. Consistent hashing
 B. Sharding by key prefix
 C. Randomized partitioning
 D. Static partitioning

Answer: A, C — Distributes keys more evenly across nodes.

31. Which failure scenario primarily tests network partition tolerance?
 A. Server crash
 B. Message delay across data centers
 C. Schema migration
 D. Disk failure

Answer: B — Network partitions test CAP trade-offs between consistency and availability.

32. For a geo-distributed social media platform, which strategies enhance both latency and availability?
 A. Data replication near users
 B. Asynchronous updates
 C. Strong global consistency
 D. Write-ahead logging

Answer: A, B — Replication and async writes improve performance while tolerating delays.

33. In graph databases, what happens if you replace a "friend" edge with a "follower" edge but forget to update the reverse direction?
 A. Data redundancy
 B. Loss of referential integrity
 C. Deadlock
 D. Serialization failure

Answer: B — Relationships lose bidirectional integrity.

34. Which challenges are more pronounced in NoSQL systems compared to relational databases?
 A. Lack of strict schema
 B. Weak consistency
 C. Complex joins
 D. Atomicity loss

Answer: A, B, D — NoSQL trades strictness for flexibility and scalability.

35. Which graph traversal ensures every node is visited exactly once before returning to the start?
 A. Eulerian Cycle
 B. Hamiltonian Cycle
 C. BFS
 D. Depth-limited Search

Answer: B — Hamiltonian cycles visit all vertices exactly once.

36. Which mechanisms can prevent lost updates in concurrent transactions?
 A. Two-phase locking
 B. Timestamp ordering
 C. Deferred writes
 D. Shadow paging

Answer: A, B — Both enforce serializable scheduling.

37. If an XML document needs to be transformed into HTML, which standard is most applicable?
 A. XPath
 B. XSLT
 C. XQuery
 D. XML Schema

Answer: B — XSLT transforms XML into other formats like HTML.

38. A columnar store using run-length encoding would perform best when:
 A. Column values are unique
 B. Column values are highly repetitive
 C. Query involves random writes

D. Tuple reconstruction is frequent

Answer: B — Run-length encoding compresses repeated values efficiently.

39. In distributed databases, which failures can cause partial transaction commit?
 A. Message failure
 B. Server crash
 C. Network partition
 D. Disk failure

Answer: A, B, C — Communication and node failures disrupt distributed commits.

40. Which consistency model ensures that once you've read your own write, you'll never see an older version again?
 A. Monotonic Writes
 B. Monotonic Reads
 C. Read Your Writes
 D. Causal Consistency

Answer: B — Monotonic reads guarantee non-regression in read versions.

41. A document database exposes a REST API with CRUD operations. What advantage does REST's stateless nature provide?
 A. Faster joins
 B. Easier load balancing
 C. Guaranteed consistency
 D. Fewer network calls
 Answer: B — Statelessness enables independent request handling and load distribution.

42. For a multi-tenant SaaS platform, what design patterns improve scalability?
 A. Database sharding
 B. Denormalization
 C. Data federation
 D. Centralized transaction manager

Answer: A, B, C — Distributing data and simplifying queries boosts scalability.

43. In a system emphasizing availability over consistency, what issue might arise after network partition recovery?
 A. Lost updates

B. Data divergence

C. Reduced throughput

D. Schema lock

Answer: B — Divergent replicas may exist temporarily or permanently.

44. Which concepts from UML are directly mappable to ER modeling?

A. Class ↔ Entity

B. Attribute ↔ Column

C. Association ↔ Relationship

D. Method ↔ Constraint

Answer: A, B, C — UML classes/attributes/associations align with ER entities.

45. What is the primary benefit of using a buffer manager in DBMS?

A. Query parsing

B. Minimizing disk I/O

C. Enforcing foreign keys

D. Handling deadlocks

Answer: B — Buffer managers reduce disk access by caching frequently used pages.

46. To guarantee atomic commit in distributed transactions, which protocols are essential?

A. Two-Phase Commit

B. Gossip Protocol

C. Write-Ahead Logging

D. Consensus Algorithms (e.g., Paxos)

Answer: A, D — Both ensure agreement on commit across nodes.

47. Why might homogeneous structure in relational tables be a limitation?

A. Difficult to normalize

B. Restricts hierarchical data representation

C. Increases redundancy

D. Slows indexing

Answer: B — Fixed schema restricts nested or irregular data modeling.

48. A property graph can represent multiple edge types such as "works_with" and "reports_to." What does this achieve?

A. Schema rigidity
B. Semantic richness
C. Easier traversal
D. Query optimization

Answer: B, C — Multiple relation types enrich semantics and improve navigation.

49. During map phase of MapReduce, if keys are not properly partitioned, what could happen?
A. Incorrect aggregation
B. Shuffle phase failure
C. Skewed reducer load
D. Network congestion reduction

Answer: C — Poor key partitioning creates load imbalance.

50. A distributed document database requires strong consistency for financial data and high availability for logs. How should the architecture handle this?
A. Use CAP trade-offs per data type
B. Replicate financial data synchronously
C. Use eventual consistency for logs
D. Force global ACID compliance

Answer: A, B, C — Hybrid consistency models satisfy different workloads.