Data Engineering & ETL Quiz Instructions: Select the best answer(s). Multi-select where noted.

(Multi-select) Which mechanisms help achieve exactly-once semantics in a Kafka + Spark Stream- ing pipeline?
A) Idempotent writes at the sink
B) At-least-once message delivery
C) Checkpointing with offset tracking
D) Stateless transformations

In a star schema, the large transactional table that stores numeric measures is called the:
A) Fact table
B) Dimension table
C) Bridge table
D) Lookup table

(Multi-select) Delta Lake provides which ACID guarantees by default?
A) Atomicity
B) Consistency
C) Isolation
D) Durability

Which scheduling strategy in Apache Airflow is best for pipelines triggered by external file arrival?
A) Cron schedule
B) Trigger rule
C) Sensor operator
D) DAG concurrency control

ELT differs from traditional ETL primarily because:
A) Transformations occur after data is loaded into the warehouse
B) It always uses batch jobs
C) It requires proprietary appliances

D) It eliminates data-validation steps

In columnar storage formats like Parquet, predicate push-down improves query performance by:
A) Reading only relevant row groups
B) Repartitioning files on the fly
C) Compressing the entire dataset
D) Increasing the replication factor

The most appropriate way to version schema changes in a data lake is:
A) Keep separate folders per date
B) Use Delta Lake time-travel
C) Rename columns directly in-place
D) Drop and recreate tables