Q51. Design data validation checks.

I would validate schema, ranges, nulls, and referential integrity at each stage.

Q52. Design data freshness tracking.

I would monitor max event timestamps and alert on SLA breaches.

Q53. Design completeness monitoring.

I would compare expected versus actual record counts.

Q54. Design data quality metrics.

I would track freshness, completeness, accuracy, and consistency.

Q55. Design schema enforcement.

I would reject or quarantine invalid records early.

Q56. Design data contract system.

I would define producer-consumer contracts with versioning.

Q57. Handle breaking schema changes.

I would block incompatible changes and require coordinated rollout.

Q58. Design PII handling system.

I would mask or tokenize sensitive fields and restrict access.

Q59. Design data lineage tracking.

I would capture metadata on data movement and transformations.

Q60. Design audit and compliance pipelines.

I would log access and changes for compliance audits.

Q61. Design cost-efficient pipeline.

I would process incrementally and avoid full recomputation.

Q62. Optimize storage cost.

I would compress data and apply lifecycle rules.

Q63. Optimize compute cost.

I would right-size clusters and use autoscaling.

Q64. Minimize full table scans.

I would partition and cluster data effectively.

Q65. Design retention and archival.

I would archive old data to cheaper storage.

Q66. Control runaway queries.

I would enforce query limits and timeouts.

Q67. Detect wasted compute.

I would monitor idle resources and optimize usage.

Q68. Design tiered storage.

I would move data across tiers based on access patterns.

Q69. Balance freshness vs cost.

I would align SLAs with business needs.

Q70. Design chargeback system.

I would track usage per team and allocate costs.

Q71. Design pipeline monitoring.

I would expose metrics and alerts.

Q72. Design SLA tracking.

I would define and monitor SLAs per dataset.

Q73. Design observability.

I would combine logs, metrics, and traces.

Q74. Detect pipeline stalls.

I would monitor lag and inactivity.

Q75. Design lag monitoring.

I would track Kafka consumer lag and processing delays.