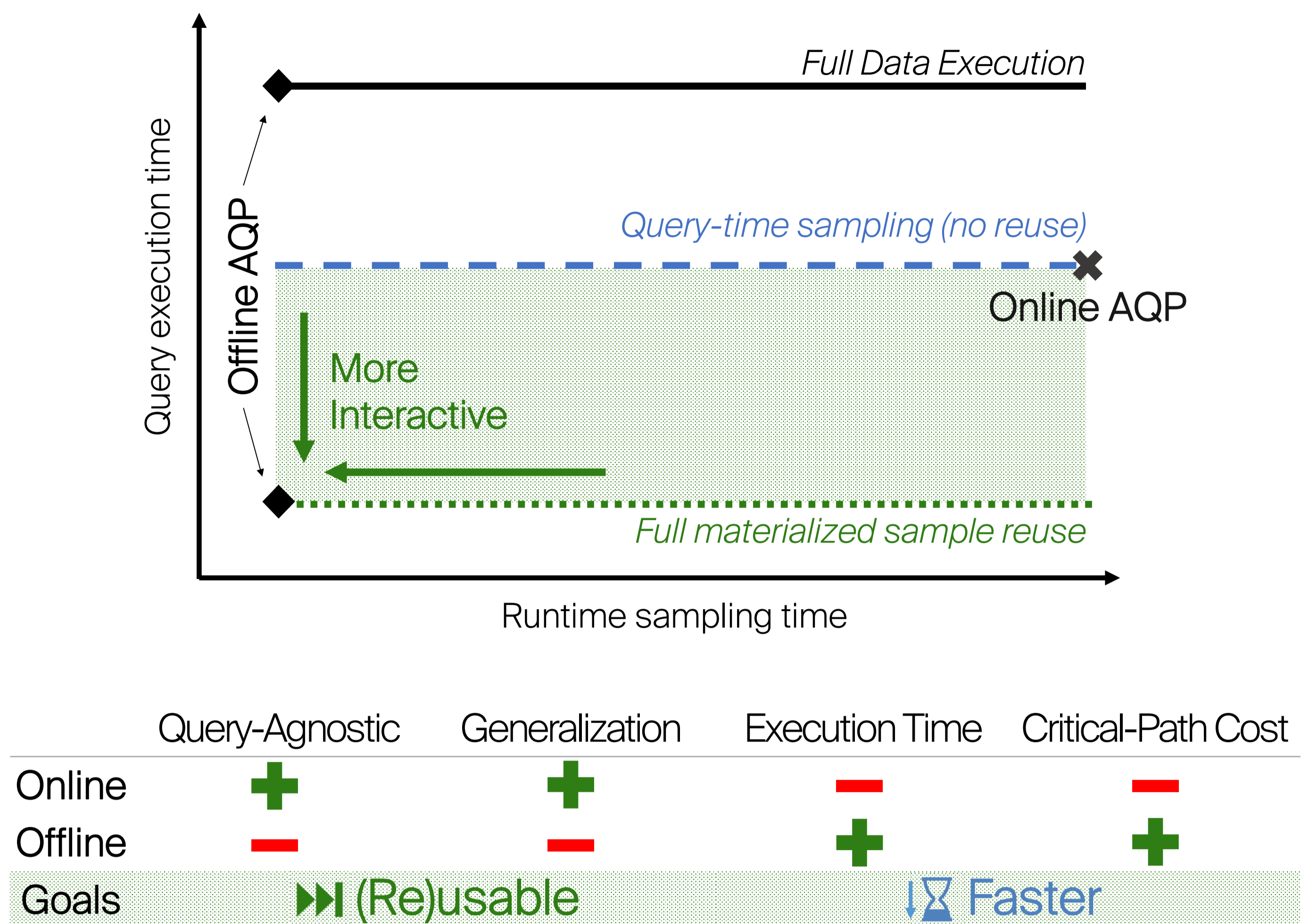


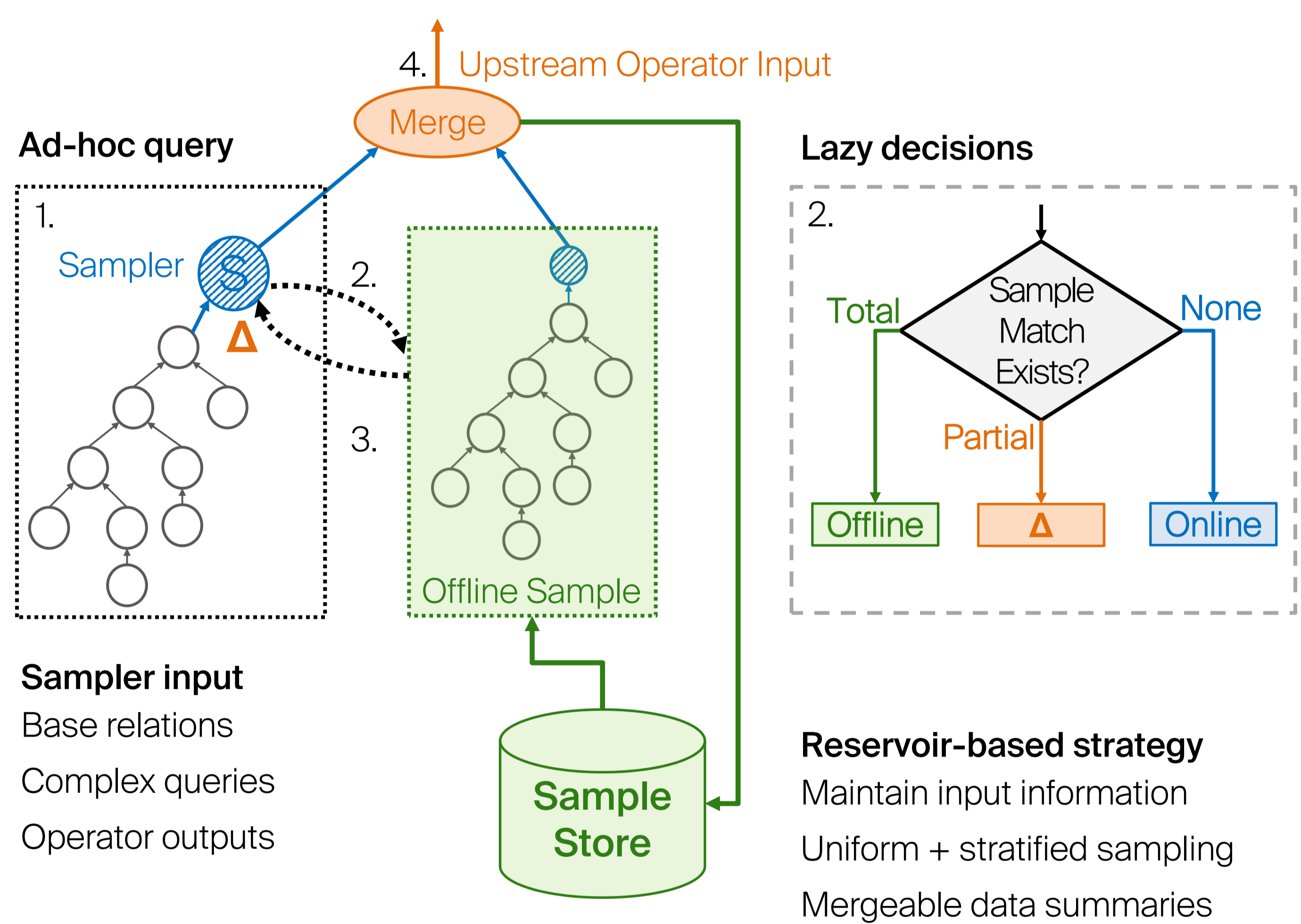
LAQy: Efficient and Reusable Query Approximations via Lazy Sampling

Strict Dichotomy of Sampling Approaches



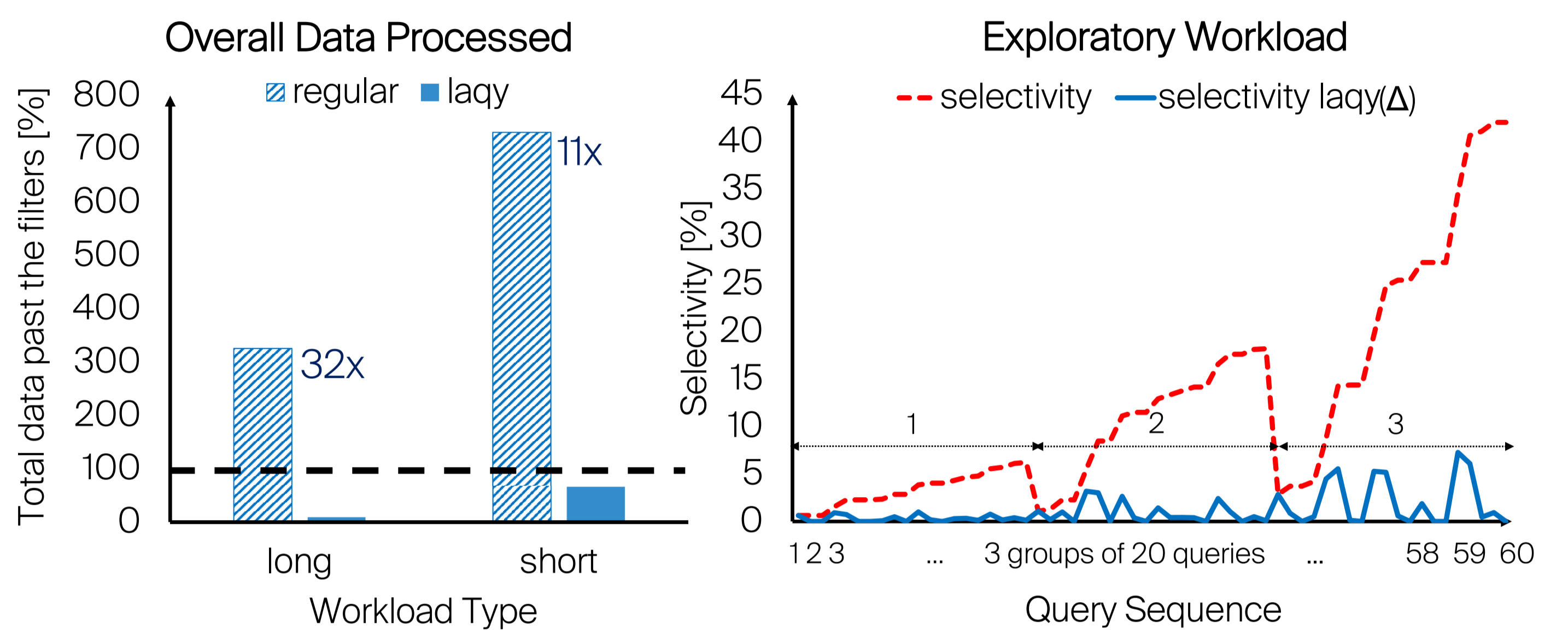
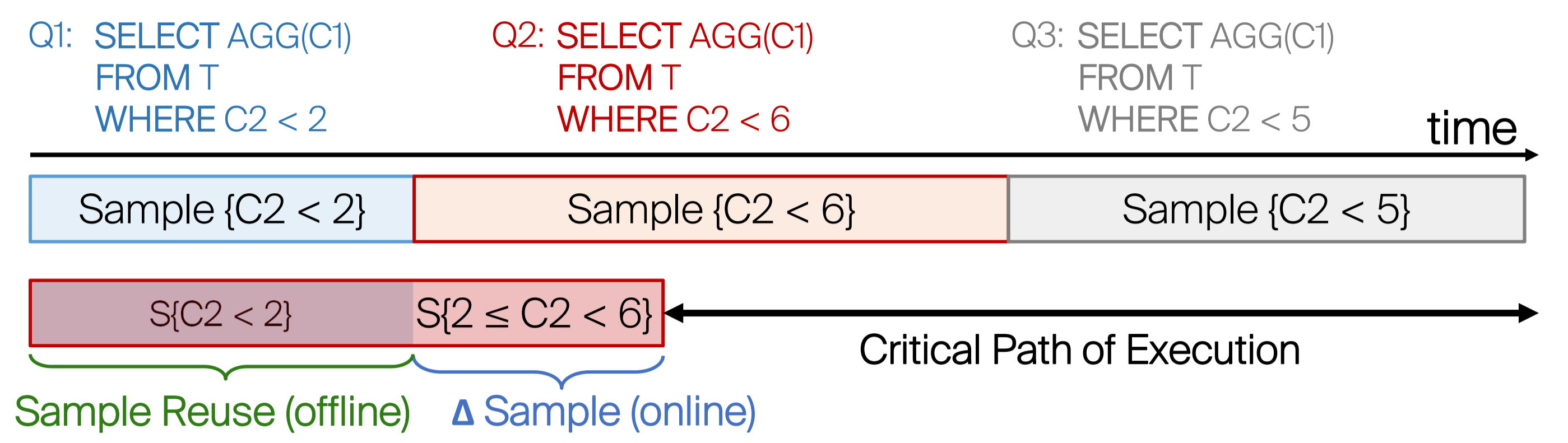
Efficient sampling = online flexibility + offline performance

LAQy: Lazy Approximate Querying



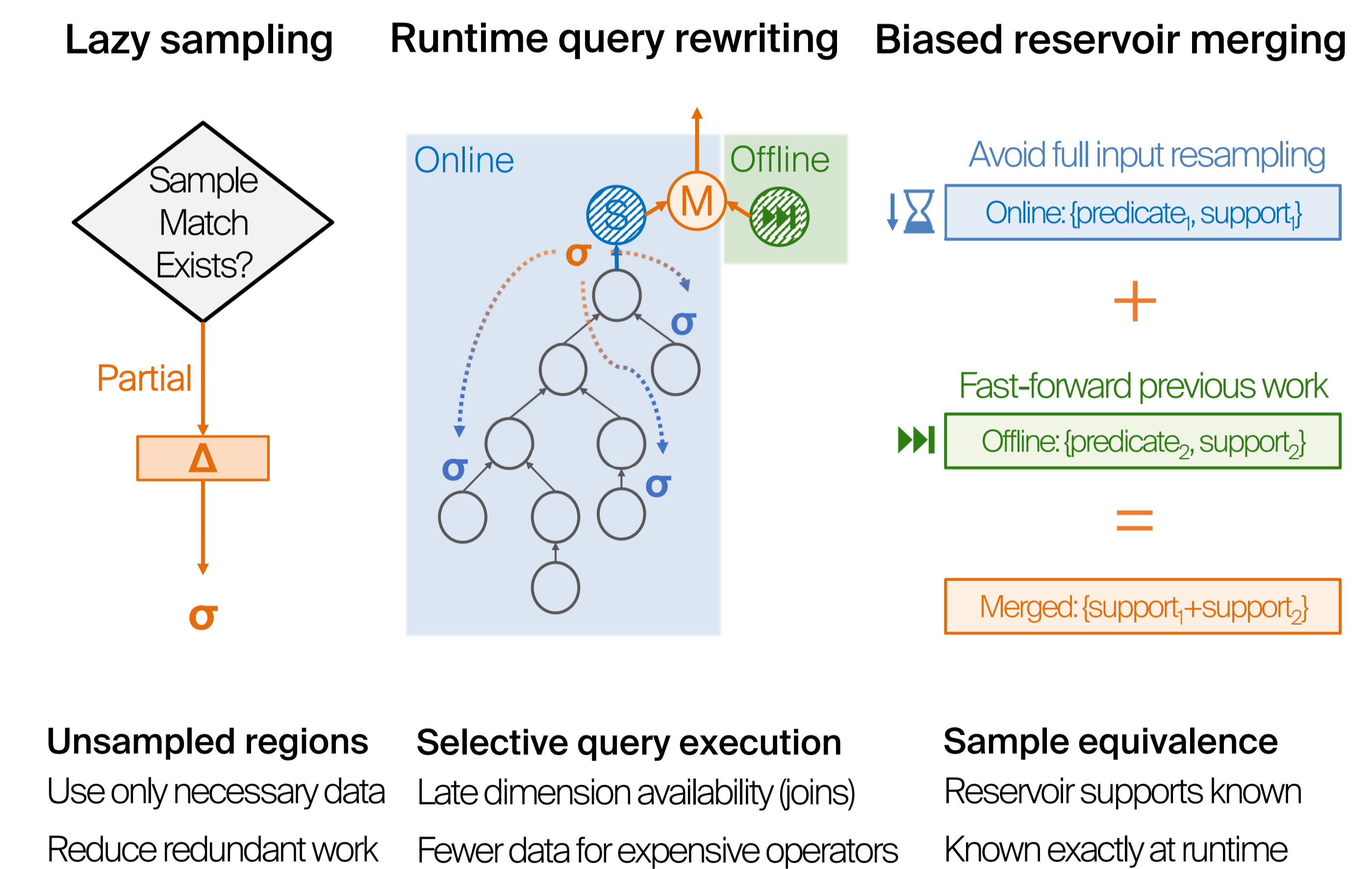
Flexible work reuse with judicious online sampling

Online Sampling is Adaptive but Workload-Oblivious



Adaptivity at the cost of avoidable critical overhead

Delta Sampling and Partial Query Execution

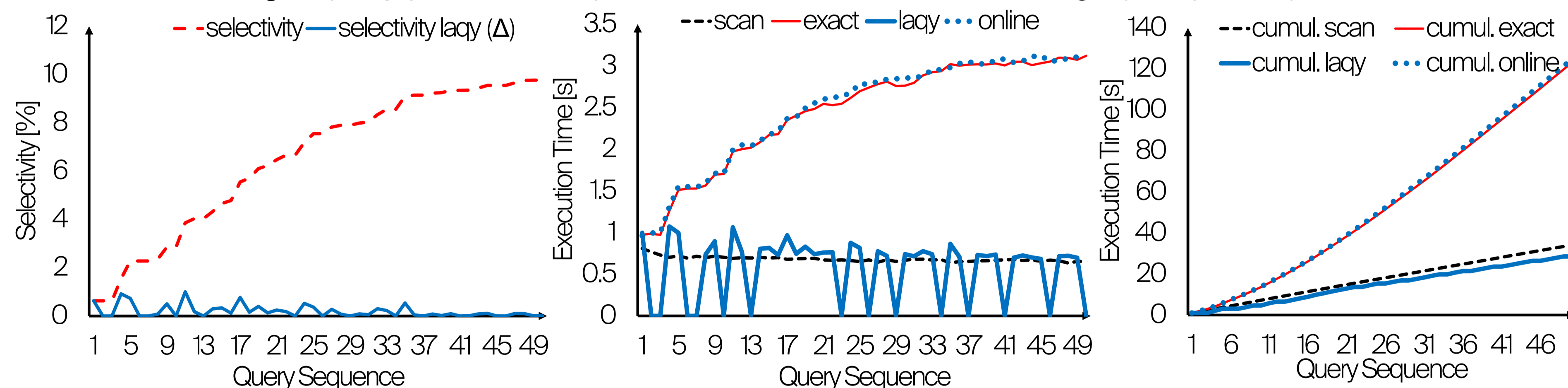


Fast and selective sample creation and merging

Exploiting Reuse Opportunities in Exploratory Workloads with Modern Analytical Systems

Setup: dual-socket Intel Xeon Gold 5118 (2x12 cores, 48 threads), 384GB RAM Data: SSB with 6B (SF1000) tuples in a fact table, a column has ~23GB
Workload: simulated exploratory workload, probabilistic change of predicates for exploration patterns Query: scan → filter → join → group by/stratification

Slow focus change: query pattern with predicate and stratification change (60 queries)



Lazy sampling is practical in modern analytics
Fast & efficient for in-memory scale-up systems

Workload awareness reduces runtime overhead
Reuse spectrum of online and offline sampling

Get LAQy! Adaptive and nimble runtime sampling
Selectivity-induced speedup + work skipping

Fast focus change: 3 query patterns with predicate and stratification change (3x20 queries)

