LAQy: Efficient and Reusable Query Approximations via Lazy Sampling

Viktor Sanca Periklis Chrysogelos Anastasia Ailamaki

Offline

Delta Sampling and Partial Query Execution

Lazy decisions

Ad-hoc query

Sampler input

Base relations

Complex queries

Operator outputs

Sample Store

Reservoir-based strategy

Maintain input information

Uniform + stratified sampling

Mergeable data summaries

Flexible work reuse with judicious online sampling

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 (SP1000) 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)

LAQy Online Exact

Strict Dichotomy of Sampling Approaches

Full Data Execution

Online AQP

Offline AQP

Full materialized sample reuse

LAQy: Efficient and Reusable Query Approximations via Lazy Sampling

LAQy: Lazy Approximate Querying

Efficient sampling = online flexibility + offline performance

Critical Path of Execution

Overall Data Processed

45

30

20

10

Selectivity [%]

Online

Offline

Goals

(Re)usable

Faster

Adaptivity at the cost of avoidable critical overhead

Explore reuse opportunities in exploratory workloads with modern analytical systems

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