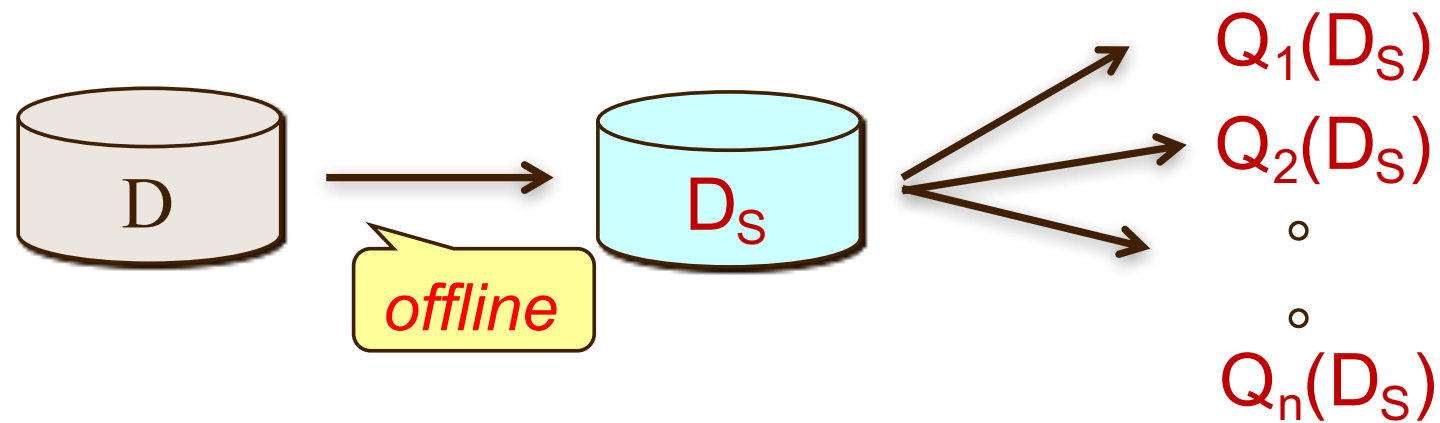


Approximate Query Processing

► Synopsis-based approximation

- for any database D compute an *one-size-fit-all* synopsis D_s
- for all queries Q posted on D , compute $Q(D_s)$ as the answer



- aggregate query only
- assumption: query load or predicates are known in advance
- *probabilistic accuracy bound: can't trust any answer in $Q(D_s)$*

► Online sampling

- no (or very bad probabilistic) accuracy bound (aggregate queries)

Bounded Approximation Scheme

Input: A resource ratio $\alpha \in (0, 1]$ and an access schema \mathcal{A} .

Scheme^[6]: Given generic SQL Q and D , computes $(Q(D_Q), \eta)$:

- ▶ access a fraction D_Q of D with $|D_Q| \leq \alpha |D|$;
- ▶ $\text{accuracy}(Q, D, Q(D_Q)) \geq \eta$.

