Sample Definition
- declarative sample definition language
  - choice of sampling scheme is left to the system
  - based on information about expected queries and the data set itself
- parameters
  - base data (as SELECT-statement) & sample size
  - grouping columns, aggregate functions for optimization

Catalog Tables
- one logical sample may consist of multiple physical samples (e.g., stratification)
- additional system tables

Maintenance
- automatic sample maintenance
  - samples are kept up-to-date
  - immediate and deferred refresh supported
  - incremental strategies are used whenever possible

Approximate Queries
- SQL-like queries for approximate query processing
  - queries against base tables
  - distinction between existential (SOME) and statistical (APPROXIMATE) errors
- optional specification of accepted error (future work)

INTERVAL and CONFIDENCE
- computation of error bounds with user-defined parameters
  - CONFIDENCE returns the confidence for a user-defined interval
  - INTERVAL returns the interval with a user-defined confidence

Example
- implemented as additional aggregation functions
- large-sample confidence intervals

Maintenance Behind the Curtain
- intercepting DML operations
  - capture modifications of base data (INSERT, UPDATE, and DELETE)
  - compute deltas of base data
  - derive net effect on the sample
  - depending on the maintenance strategy
    - apply to sample (immediate refresh)
    - write to staging table (deferred refresh)
  - REFRESH command
    - manually refresh sample

Rewriting the Query Tree
- transparent replacement of base tables by appropriate samples
- example: existing sample over lineitem and orders (lo_sample)