

# ***LDBC SNB Benchmark Auditing***

*Moritz Kaufmann*

Technische Universität München

**Process**

**+**

**Rules**

**+**

**Metrics**

*Where to start?*

Download driver

[https://github.com/ldbc/ldbc\\_driver](https://github.com/ldbc/ldbc_driver)

Implement driver for SUT

Examples: [https://github.com/ldbc/ldbc\\_snb\\_implementations](https://github.com/ldbc/ldbc_snb_implementations)

Verify it using validation dataset + validation file

[https://github.com/ldbc-dev/ldbc\\_snb\\_interactive\\_validation](https://github.com/ldbc-dev/ldbc_snb_interactive_validation)

Download data generator + generate dataset

[https://github.com/ldbc-dev/ldbc\\_snb\\_datagen\\_0.2](https://github.com/ldbc-dev/ldbc_snb_datagen_0.2)

System with 3 years of data will be benchmarked for **2h of simulation time**

Main metric: **Throughput @ Scale**

Main obstacles: **Query latency constraints +  
Crash recovery**

**Query Implementation**

**Dataset @ SF**

**Clean Execution  
Environment**

**System under test**

**Serializability Impl. (optional)**

**1**

**Install base system**

**Setup benchmark configuration**

**Validation run**

**2**

**Reset system**

**Setup loading configuration**

**Load data (timed)**

**3**

**Setup benchmark configuration**

**Optional system warmup**

**Benchmark  $\geq$  2h of  
simulation time**



**4**

**Verify benchmark run**

**Continue run**

**Crash System +Recovery testing**

### *Additional tooling*

Validation of loaded data (potentially requires implementation by the sponsor)

Query latency checks from driver log

Serializability verification (potentially requires implementation by the sponsor)

Recovery verification

## *Report*

Throughput @ Scale

Load time

Recovery time

System configuration, cost, ...

*Inspiration from TPC, but we want to support systems without a declarative query language.*

~~Prohibit using hinted query plans, e.g. for join ordering~~

~~*Disallow precomputed indexes*~~

~~Force serializable consistency~~

### *Rough guidelines*

Datatypes have to conform to specification

Indexes have to be updatable

System has to be generally obtainable within 6 months

### *Intricate edges*

How to define the weight of each query?

How to determine cost of system?

- Cloud? Open-source?

# Discussion