METRICS FOR MEASURING THE PERFORMANCE OF THE MIXED WORKLOAD
CH-BENCHMARK

Florian Funke (TU München)
Alfons Kemper (TU München)
Stefan Krompass (TU München)
Thomas Neumann (TU München)
Michael Seibold (TU München)
Harumi Kuno (HP Labs)
Anisoara Nica (Sybase)
Raghunath Nambiar (Cisco)
Meikel Poess (Oracle)
AGENDA

- Motivation
- Quick Overview of TPC-C and TPC-H
- CH-BenCHmark
- Conclusion
Motivation

OnLine Transaction Processing (OLTP)

Transactional Application(s)

Transactions

Analytical Tool(s)

Queries

Business Intelligence (BI)

ETL Process

week
OnLine Transaction Processing (OLTP)

Transactions

ETL Process

day

Business Intelligence (BI)

Analytical Tool(s)

Queries

Motivation
OnLine Transaction Processing (OLTP)

Transactional Application(s)

Transactions

ETL Process

hour

Analytical Tool(s)

Queries

Business Intelligence (BI)
OnLine Transaction Processing (OLTP)

Transactional Application(s)

Transactions

Analytical Tool(s)

Queries

min

ETL Process

Business Intelligence (BI)
Hybrid OnLine Transaction Processing (OLTP) and Business Intelligence (BI)

- Transactions and Analytical queries are run against the same database
- Add hoc-queries
- Reporting queries
- System must be tuned for both

Diagram:
- Transactional Application(s)
- Analytical Tool(s)
- Transactions and Queries
- Hybrid OnLine Transaction Processing (OLTP) and Business Intelligence (BI)
TPC BENCHMARKS FOR OLTP AND DS

TpmC

Transactions

TPC-C

Similar, but not identical schema
Different scaling models
Different execution rules
Different metrics

QphH@SF

Queries

TPC-H
Some of the underlying principles of C and H are identical, but some are not.

What about:
- Schema
- DB Scaling
- Workload
- Execution Rules
- Metric
- ACID requirements
Similarities

- Model businesses that must manage, sell or distribute products or services
- Contain Orders and Customer tables
- Order-line and lineitem model sub-entities of orders

Differences

- Tables Warehouse, Stock, Neworder and District are not in TPC-H
- Tables Partsupp, Supplier, Nation and Region are not in TPC-C
CH SCHEMA

Supplier (10k) in Nation (62) in

Warehouse (W) \(\text{sup-by}\) to District (W * 10)

Stock (W * 100k) of Item (100k)

New-Order (W * 9k) available

Order-Line (W * 300k) contains

History (W * 30k)

Customer (W * 30k)

Region (5)

serves

located-in

has

issues

pending

contains
HURDLE - DB SCALING

- TPC-C employs a continuous scaling model, which causes the database to grow with system performance.
- TPC-H employs a scale factor (SF) model, where benchmark sponsors can choose the SF for a given system.
- CH uses the TPC-C scaling model:
  - Warehouse, Stock, Item, History, Neworder, Orderline, District, Customer, and Order scale according to TPC-C rules.
  - Supplier is fixed at 10,000 → an entry in the Stock table is assigned a supplier via a simple formula: \( s_{i\_id} \times s_{w\_id} \mod 10,000 = s_{suppkey} \)
  - Cardinality of Nation is increased to 62.
Mixed workload OLTP and BI

- OLTP represented with TPC-C → can be used unmodified
  - New-Order, Payment, Order-Status, Delivery, Stock-Level
  - Same mix as in TPC-C
- BI represented with TPC-H → needs to be modified
  - Queries were re-formulated to match new schema
  - Syntactical structure was preserved
  - Business semantics was preserved
EXAMPLE: QUERY 5

```
SELECT n_name, SUM(l_extendedprice * (1 - l_discount)) 
  AS revenue 
FROM customer, orders, lineitem, supplier, nation, region 
WHERE c_custkey = o_custkey 
  AND l_orderkey = o_orderkey 
  AND l_suppkey = s_suppkey 
  AND c_nationkey = s_nationkey 
  AND s_nationkey = n_nationkey 
  AND n_regionkey = r_regionkey 
  AND r_name = '[REGION]' 
  AND o_orderdate >= DATE '[DATE]' 
  AND o_orderdate < DATE '[DATE]' + INTERVAL '1' YEAR 
GROUP BY n_name ORDER BY revenue DESC
```

- Additional join to stock table
- Orderline instead of lineitem

```
SELECT n_name, SUM(ol_amount) AS revenue 
FROM customer, "order", orderline, stock, supplier, nation, region 
WHERE c_id=o_c_id AND c_w_id=o_w_id AND c_d_id=o_d_id 
  AND ol_o_id=o_id AND ol_w_id=o_w_id 
  AND ol_d_id=o_d_id 
  AND ol_w_id=s_w_id AND ol_i_id=s_i_id 
  AND mod((s_w_id * s_i_id),10000)=su_suppkey 
  AND ascii(SUBSTRING(c_state, 1, 1))=su_nationkey 
  AND su_nationkey=n_nationkey 
  AND n_regionkey=r_regionkey 
  AND r_name='[REGION]' AND o_entry_d>='[DATE]' 
GROUP BY n_name ORDER BY revenue DESC
```
Current model allows for a OLTP only, BI only or mixed workload.

Workload mix is specified as the number of OLTP and BI streams connected to the DB.

OLTP streams dispatch TPC-C transaction (according to the TPC-C mix).

BI streams each run the 22 queries in different order.

What mix is most representative?
HURDLE - METRIC

- TPC-C employs a throughput metric \([\text{TpmC}]\)
- TPC-H employs a geometric mean of both a response time and throughput metric
  - Simple model could be to report:
    - Transaction Throughput \([\text{TpmC}]\)
    - Analytical Query Throughput \([\text{QphH}]\)
- Higher transactional throughput may result in larger data volume which in turn may result in longer response times for analytical queries
- Idea is to monitor data volume growth and normalize \(\text{QphH}\) accordingly
Presented CH BenCHmark, a benchmark that models both OLTP and BI workloads

A CH like benchmark is needed to analyze systems that are capable of running mixed workloads

Based on TPC-C and TPC-H

Most of the work for such a benchmark is completed:
- Schema and scaling rules
- Data generator modifications
- Queries
- Execution rules

What is missing: Specification
MOTIVATION

Transactional Application(s)

Transactions

OnLine Transaction Processing (OLTP)

ETL Process

Analytical Tool(s)

Queries

Decision Support (DS)
OnLine Transaction Processing (OLTP)

ETL Process

Decision Support (DS)

Transactions

Queries

Transactional Application(s)

Analytical Tool(s)
CH-BENCHMARK

- **Schema**
  - Unmodified TPC-C schema
  - Added Supplier and Nation tables from TPC-H

- **Workload**
  - TPC-C transactions run without modifications
    - New-Order, Payment, Order-Status, Delivery, Stock-Level
  - Modified TPC-H queries to match the TPC-C schema
    - Same syntactical structure as TPC-H
    - Same business semantics as TPC-H

- **Scaling**
  - Scaling model from TPC-C
    - Warehouse, Stock, Item, History, Neworder, Orderline, District, Customer, and Order scale according to the TPC-C rules
    - Supplier is populated with fixed number (10,000)