



IBDB: A Benchmark Suite for Industrial Big Data System

Yijian Cheng



AGENDA

BACKGROUND

ARCHITECTURE

INSTRUCTIONS

EXPERIMENTS

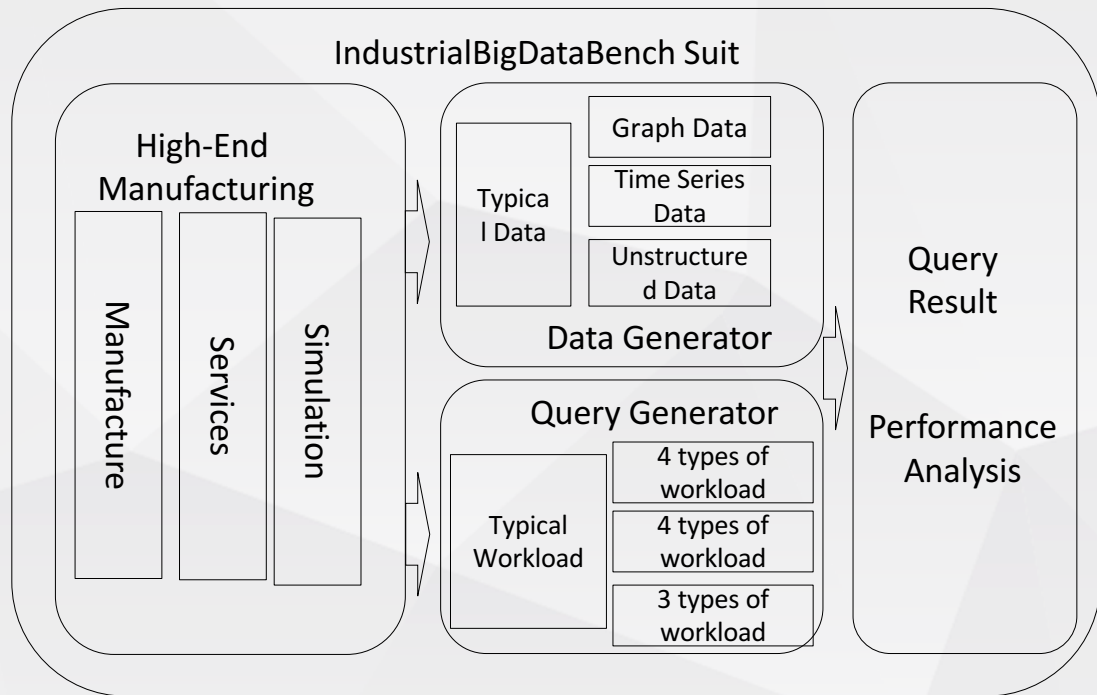


BACKGROUND

- ❑ IoT, Intelligent Manufacture, High-End Manufacture
- ❑ Increasing industrial data over decades
- ❑ The diversity of industrial applications
- ❑ The complexity of datasets and workload



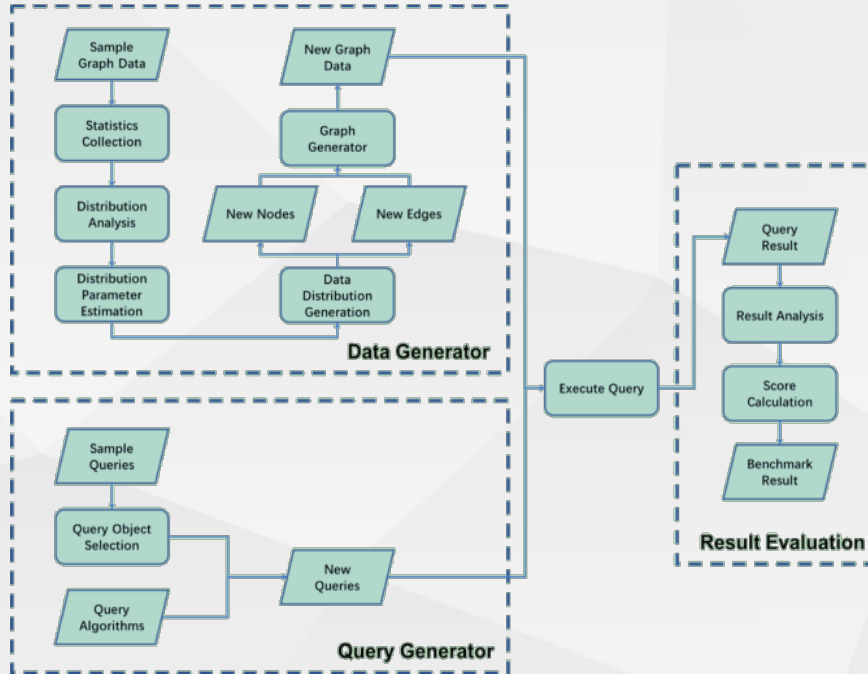
ARCHITECTURE



IndustrialBigDataBench :

- Graph-benchmark
 - ✓ Generator
 - ✓ Workload
 - ✓ Metrics
- Time Series-benchmark
 - ✓ Generator
 - ✓ Workload
 - ✓ Metrics
- Unstructured-benchmark
 - ✓ Generator
 - ✓ Workload
 - ✓ Metrics

GRAPH DB-BENCHMARK



❑ Data Generator Module

✓ Generate datasets by scaling the sample data.

❑ Query Generator Module

✓ Generate the four types of workloads

❑ Result Evaluation Module

✓ Generate script to execute the workloads and visualize the result



GRAPH DB-BENCHMARK

Scenario

✓ BOM data management in high-end manufacture

Data Generator

✓ **Data analysis:** Analysis of the distribution characteristics of sample BoM map data

✓ **Scalable data generator:** generate BOM data by scaling the sample data

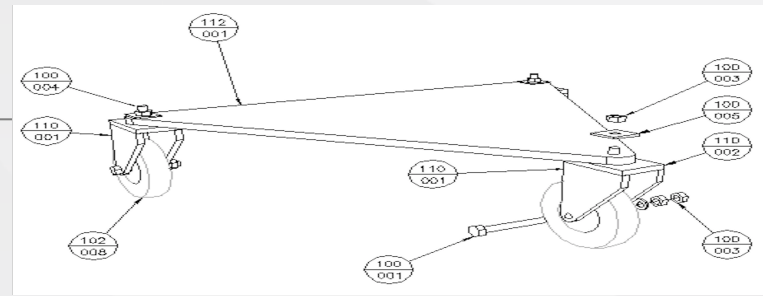
Workload&Metrics

✓ **where used** is used to search for the usage of a part in the BoM, response time(ms)

✓ **generate structure** is to used to generate the overall structure of a product, response time(ms)

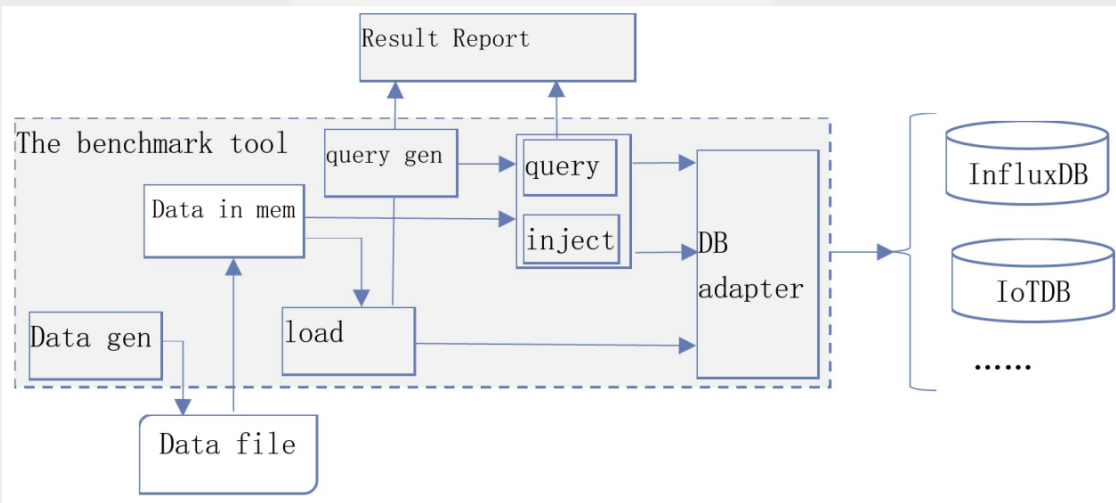
✓ **structure diff** is used to compare the differences between the two structures, response time(ms)

✓ **structure aggr** is to used to obtains purchase plan and part borrowing lists, response time(ms)





Time Series DB-BENCHMARK



❑ DataGen Module

- ✓ train data model and generate datasets

❑ DataLoad Module

- ✓ test the performance of data loading

❑ QueryGen Module

- ✓ generate workload and perform the test

❑ ResultReport Module

- ✓ save the result and visualize the result

❑ DBAdapter Module

- ✓ Shield differences between systems



Time Series DB-BENCHMARK

❑ Scenario

- ✓ **Wind Turbines Monitoring** : Each wind turbine has several sensors, which collect data from the wind turbine itself and environmental data.

❑ Data Generator

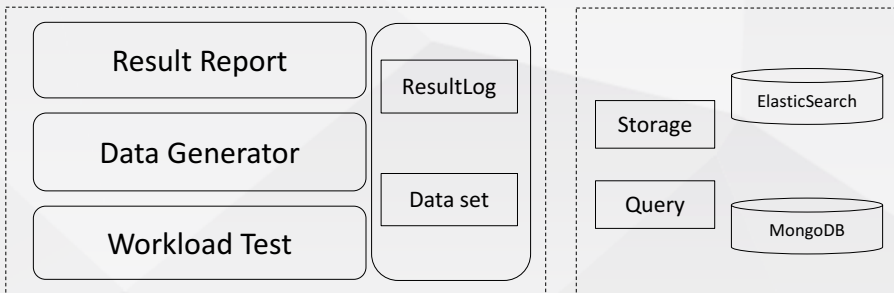
- ✓ **Wind Data**: use the real datasets provided by a company to train the wind model; simulate the power generation according to the cut-in and cut-out of the fan, the energy conversion mechanism.
- ✓ **Environmental data**: Generate data using ARIMA models

❑ Workload&Metrics

- ✓ **Load** : generates sensor data for one month, then tests the import performance of the target database, points/s
- ✓ **Append** : continuously increases the number of devices to append data to the target database, points/s
- ✓ **Read and Analysis** : Simple window range queries and aggregated queries by time, requests/s, avg response time(us)
- ✓ **Stress Test** : Increase the number of query while appending data and increase the number of devices to append data while querying. (requests/sec, average response time(us) and points/sec)



UNSTRUCTURED DB-BENCHMARK



DataGen Module

- ✓ Set the parameter of the generation
- ✓ Generate simulation file and binary file

Workload Test Module

- ✓ Simulate the most common file operation

ResultReport Module

- ✓ save the result and visualize the result



UNSTRUCTURED DB-BENCHMARK

Scenario:

- ✓ Simulation file management

Data Generator

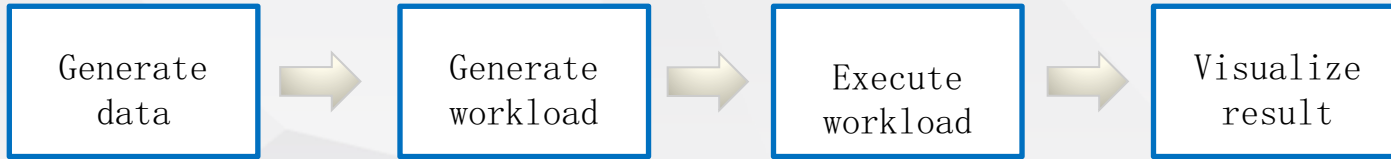
- ✓ We generate the simulation files according to the type, size and quantity of the real simulation files.

Workload&Metrics

- ✓ **Data transmission:** includes data generation and file uploading.
- ✓ **Simple query:** includes detection on the uploaded file, annotating the exception file, analyzing and calculating the statistical value of the exception file.
- ✓ **complex query:** includes version management test, abnormal file detection test, hot-file query test and other load tests with features. It can be used to meet users' evaluation requirements in different levels.



INSTRUCTIONS



Step1: generate datasets according to the data models and the parameters you specified

Step2: generate the workloads according to the specific domains

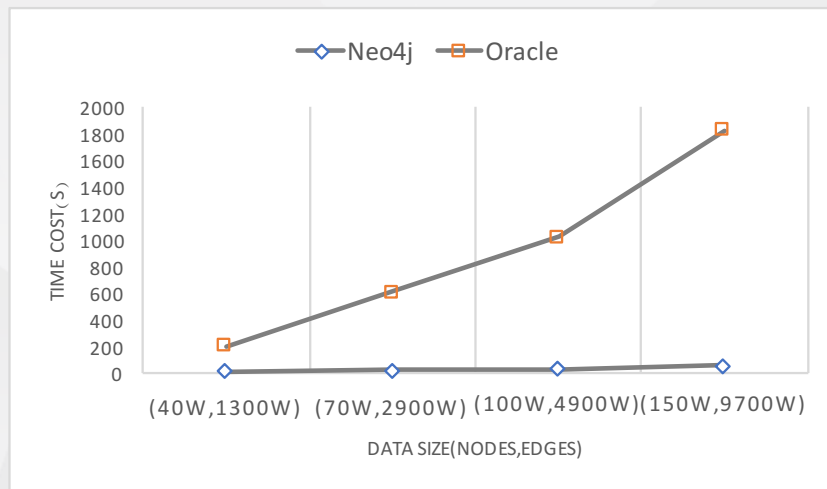
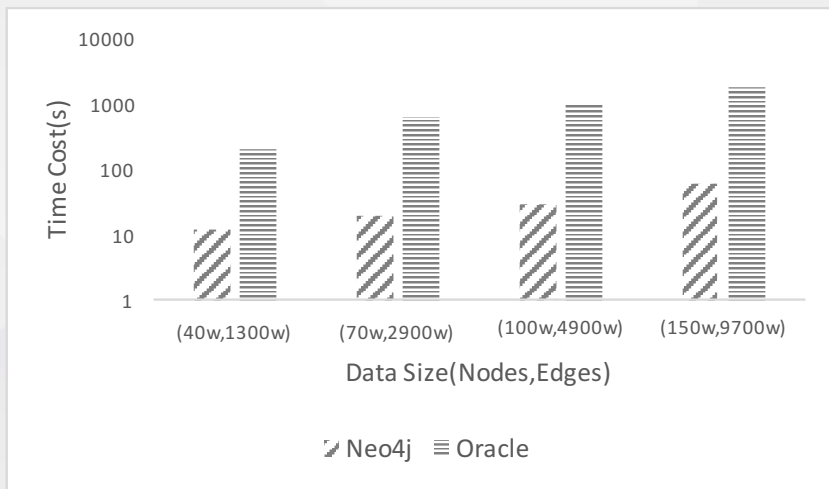
Step3: execute the workloads

Step4: visualize the results



EXPERIMENTS—Graph DB-benchmark

➤ Import test

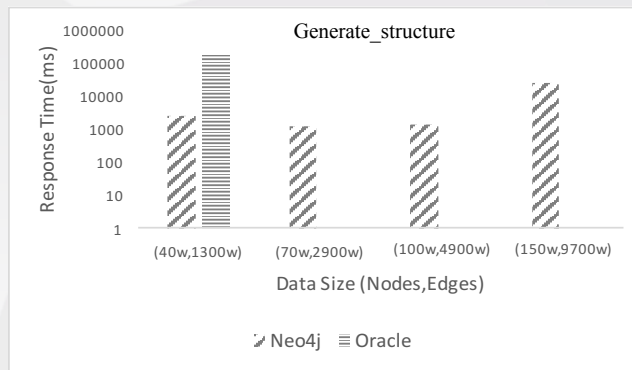
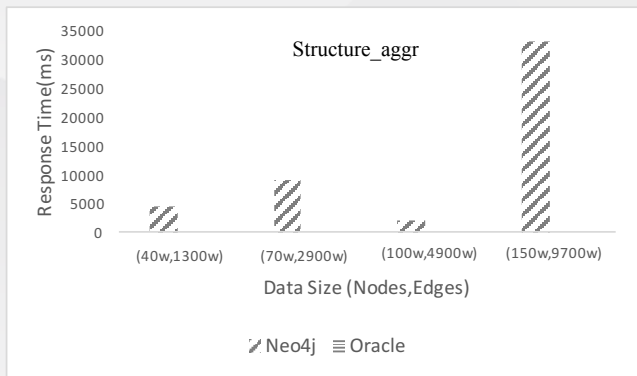
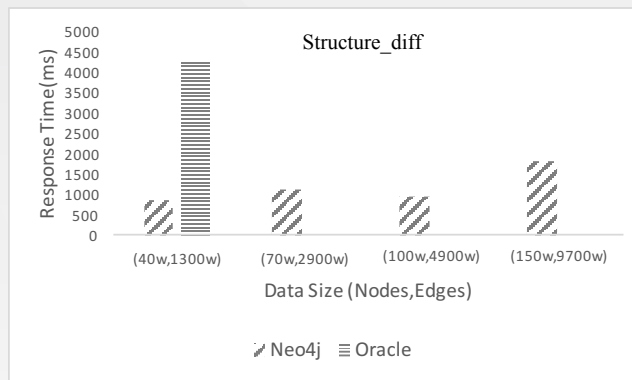
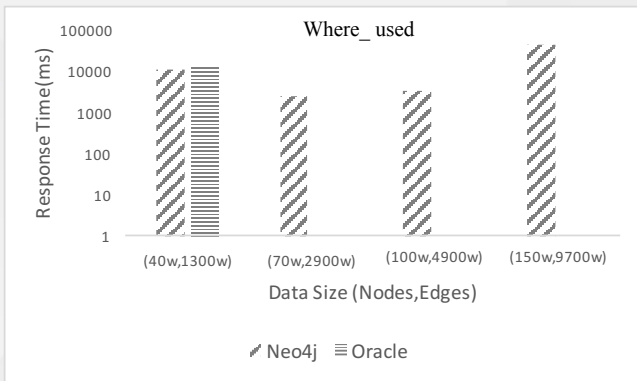


Performance Comparison between Neo4j and Oracle



EXPERIMENTS—Graph DB-benchmark

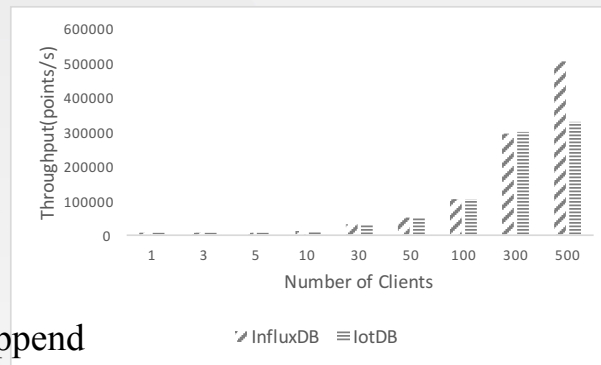
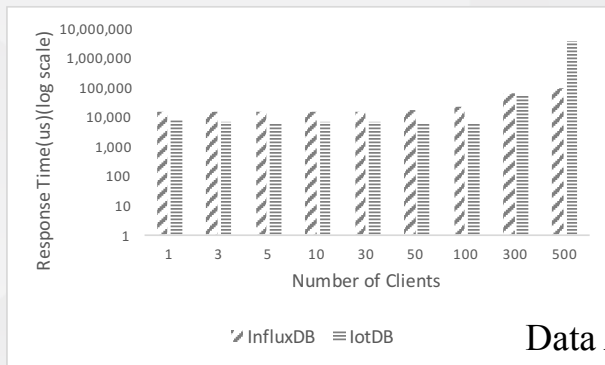
➤ Performance Comparison for Query between Neo4j and Oracle



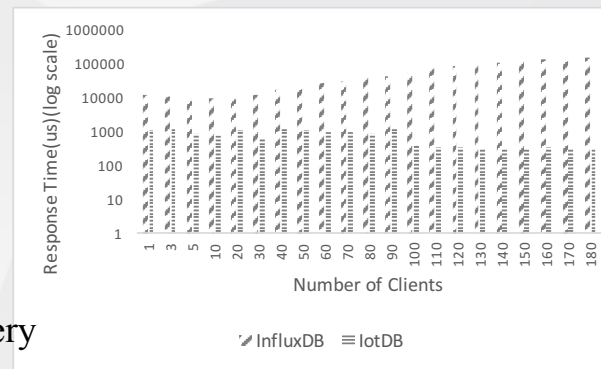
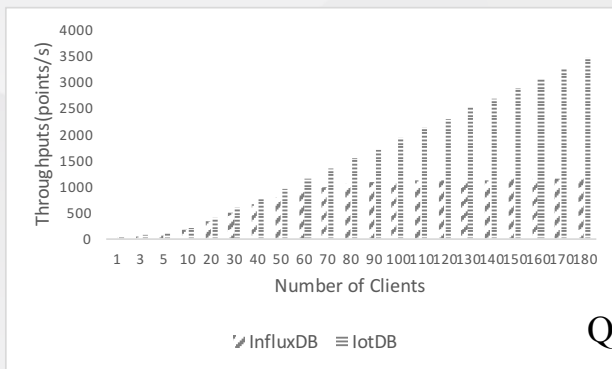


EXPERIMENTS—Time Series DB-benchmark

Response Time and Throughput comparison for Data Append and Query



Data Append

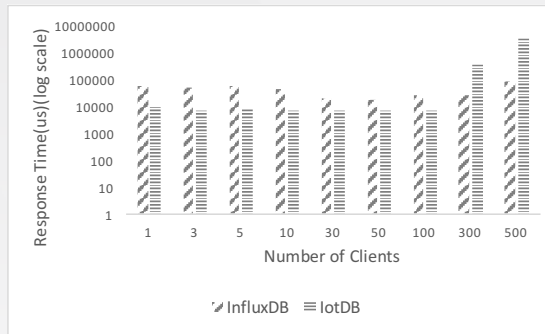
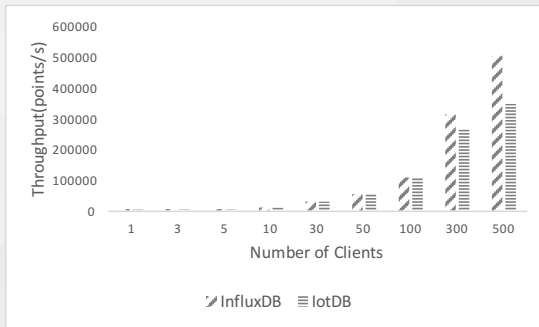


Query

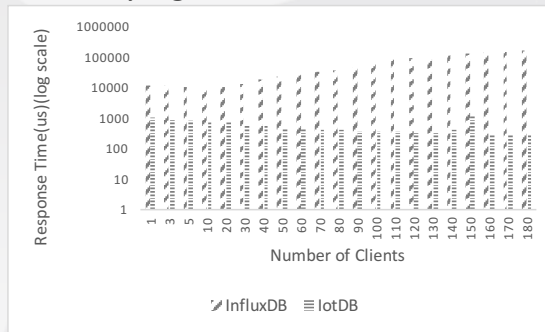
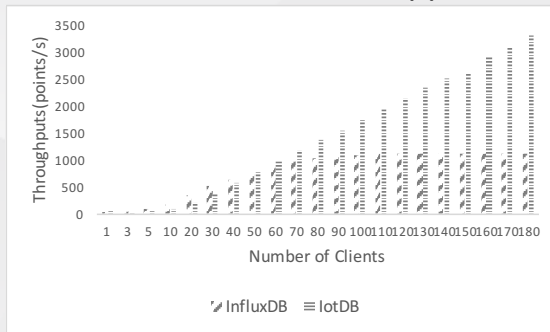


EXPERIMENTS—Time Series DB-benchmark

Performance Comparison for Stress Testing between InfluxDB and lotDB



Append While Querying

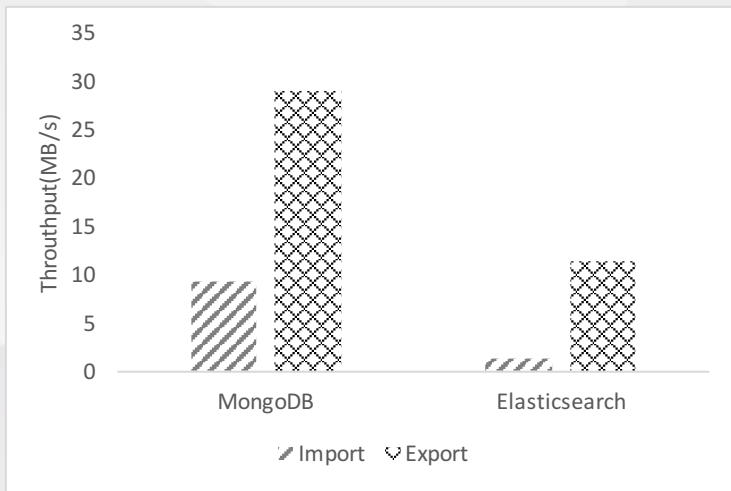


Query While Appending

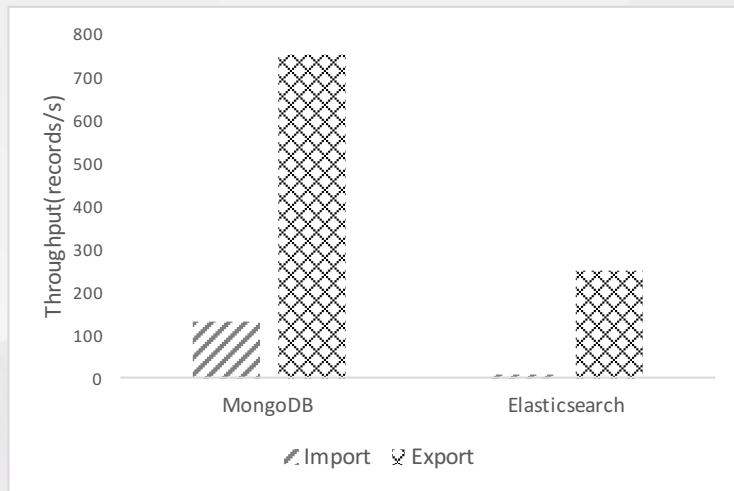


EXPERIMENTS—Unstructured DB-benchmark

Performance Comparison for Import and Export



Binary File

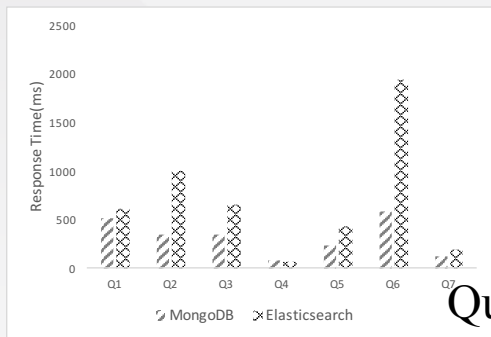


Metadata

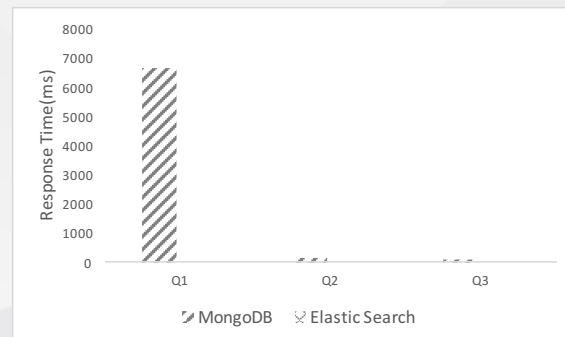
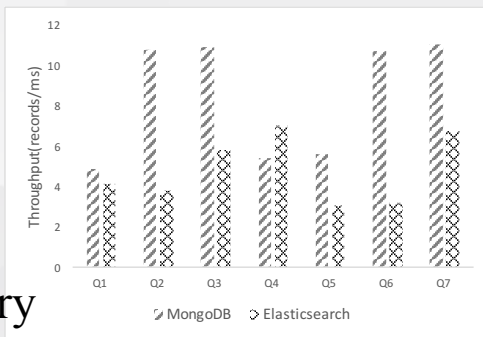


EXPERIMENTS—Unstructured DB-benchmark

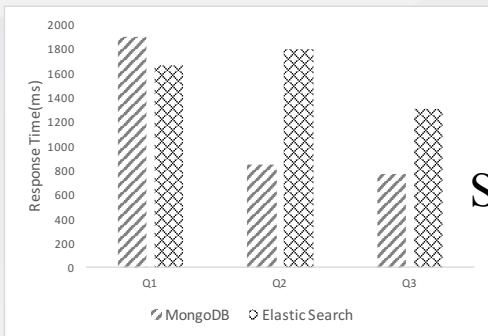
Performance Comparison for Query, Sort and Statistics



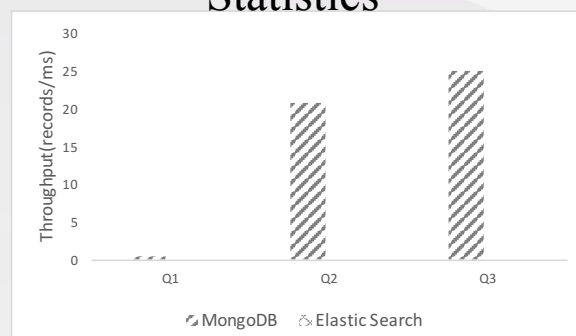
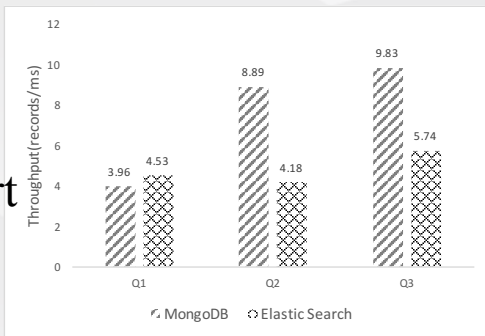
Query



Statistics



Sort





Thank you!