Apache Solr Performance Test on Range & Substring queries

1.0 Introduction

To test the performance in the context of range and substring queries in Apache Solr following configurations were used.

- 5 types of query tests
    - Date range queries
    - Energy range queries
    - Data and energy range queries
    - Substring queries
    - Substring, data and energy range queries
- 3 varying number of records
    - 100,000 records (Solr data size – 136MB)
    - 1,000,000 records (Solr data size – 1.27GB)
    - 10,000,000 records (Solr data size – 12.11GB)

Gathered data is represented using 3 types of graphs

1. Average time per record (total)vs number of records
   Total average time of all the queries for the corresponding query type
2. Average time per query vs number of records
3. Average time per record (grouped by query type) vs number of records

All the query input values for the parameters were randomly generated.

- For example for a Date range [date TO date], the dates are randomly generated so that all the 1000 queries of the type would be with different input values.

2.0 Data representation

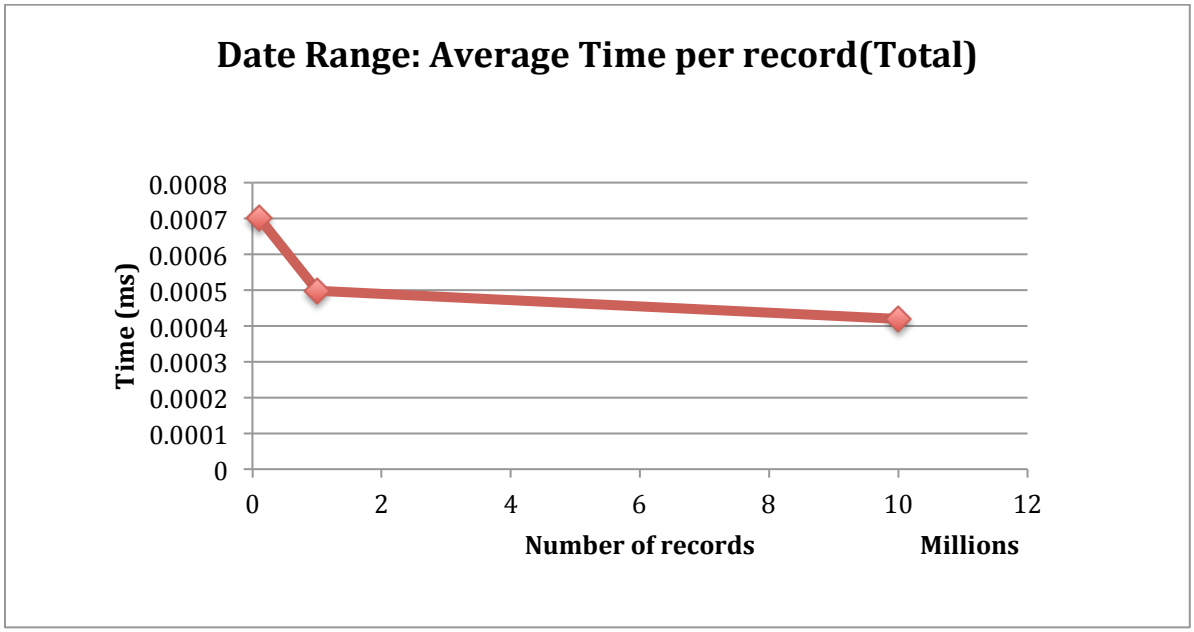
2.1 Range queries on the Date(3000 Queries)

Executed queries:

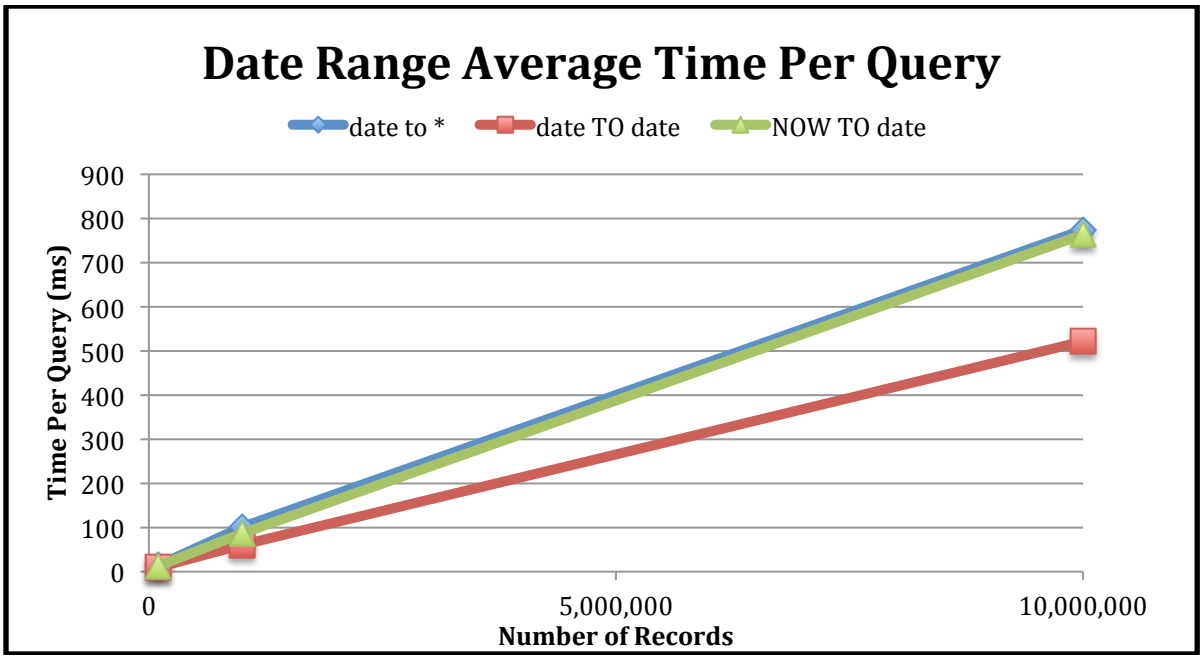
- createdDate:[date TO *] : 1000 queries
- createdDate:[date TO date] : 1000 queries
- createdDate:[NOW TO date] : 1000 queries

Each of the 3 types of queries were run 1000 times with randomly generated dates.

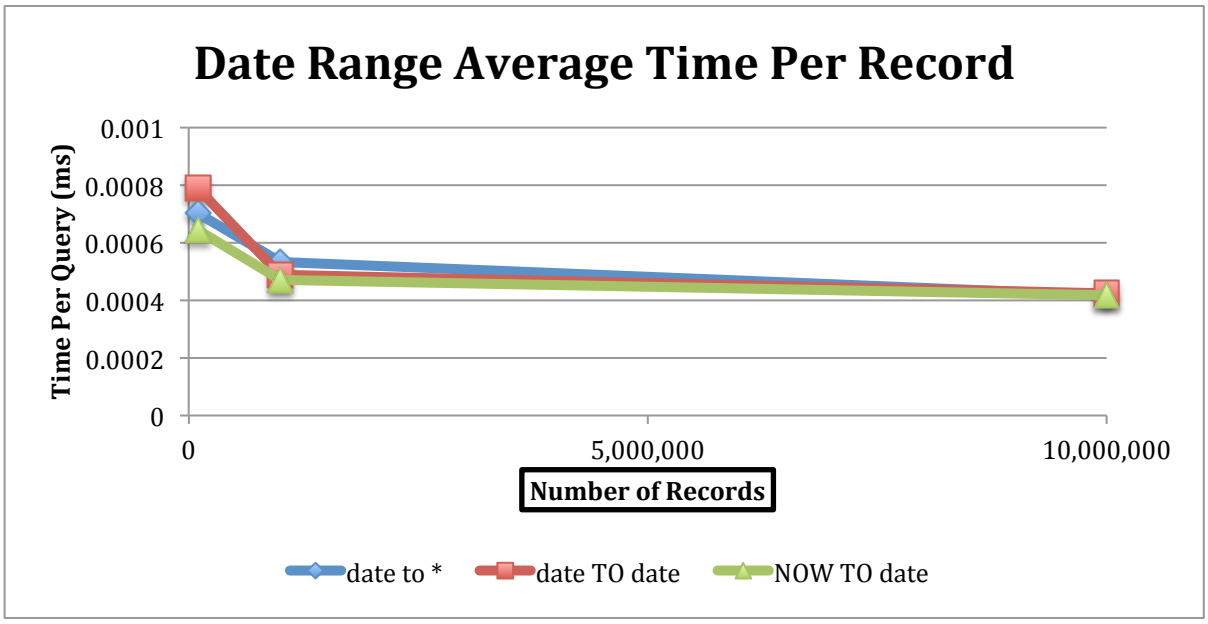
2.1.1 Average time per record (total) vs number of records



2.1.2 Average time per query vs Number of records



2.1.3 Average time per records (grouped by query type) vs Number of records



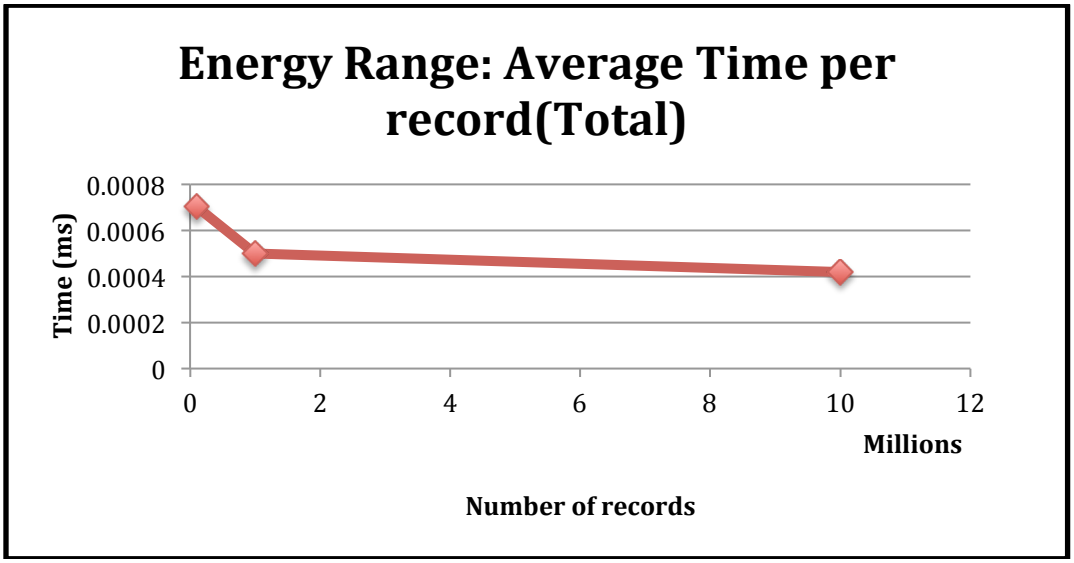
2.2 Energy Range queries (3000 Queries)

Executed queries:

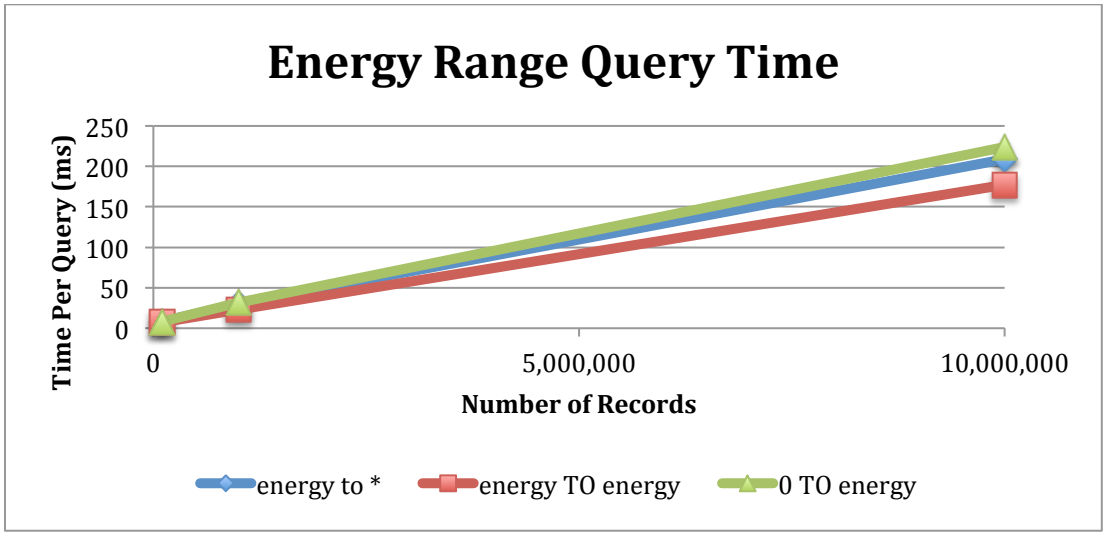
- finalEnergy:[energy TO *] : 1000 queries
- finalEnergy:[energy TO energy] : 1000 queries
- finalEnergy:[0 TO date] : 1000 queries

Each of the 3 types of queries were run 1000 times with randomly generated energy values.

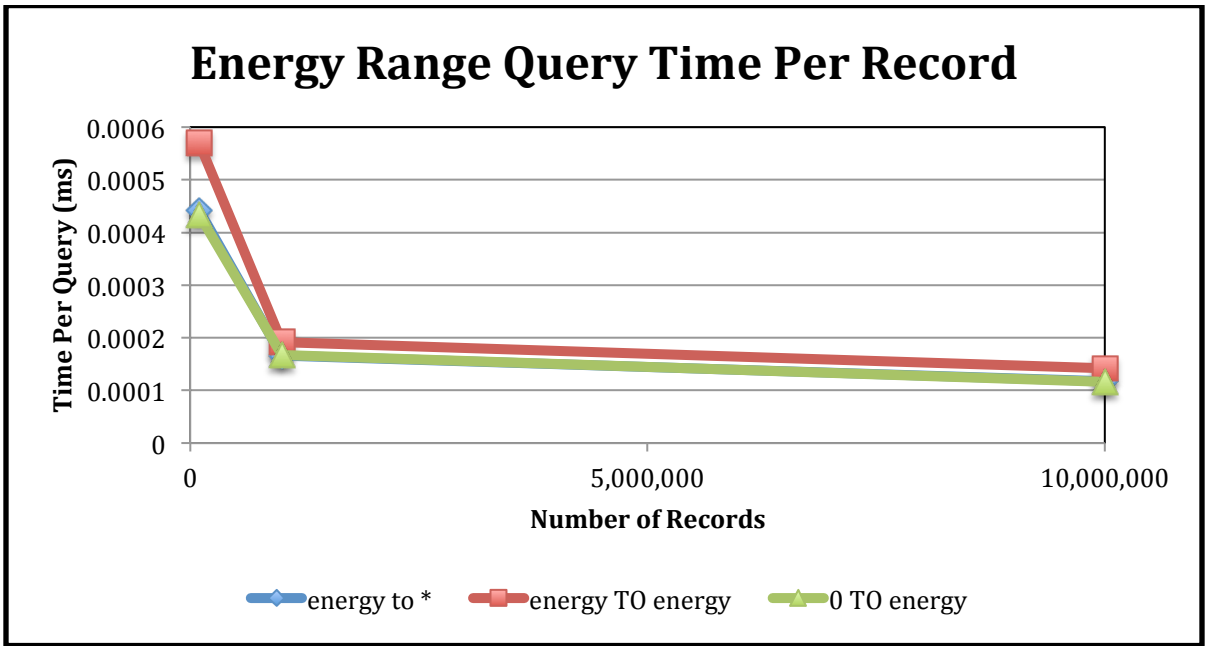
2.2.1 Average time per record (total) vs number of records



2.2.2 Average time per query vs Number of records



2.2.3 Average time per records (grouped by query type) vs Number of records



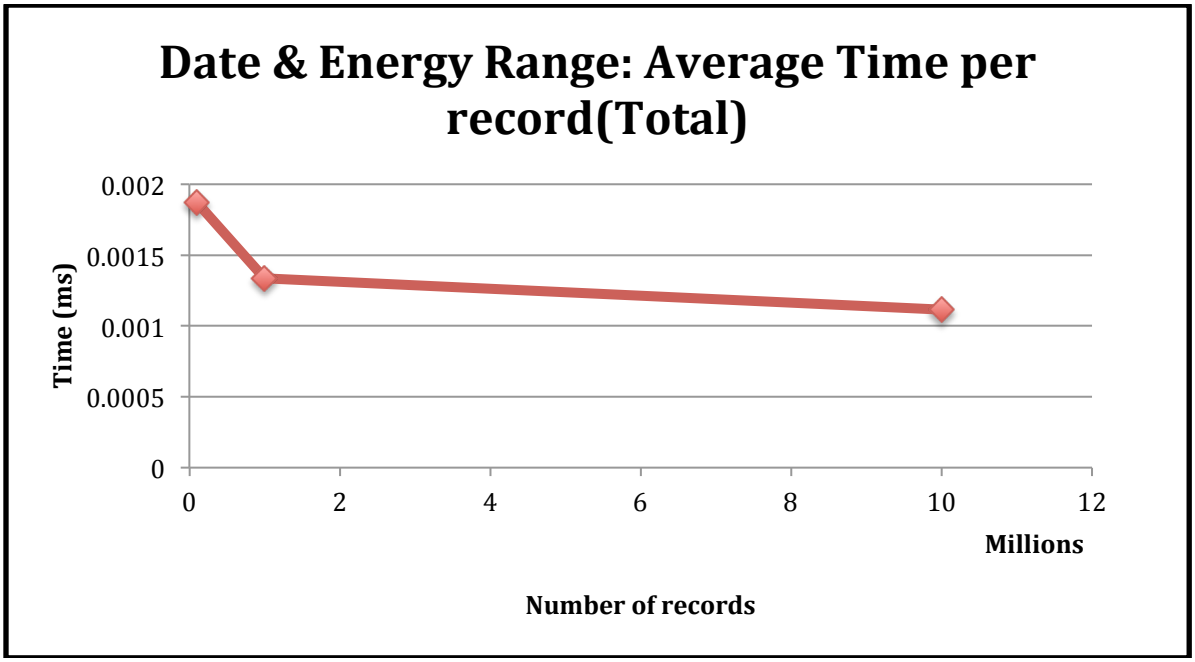
2.3 Data and Energy Range queries (3000 Queries)

Executed queries:

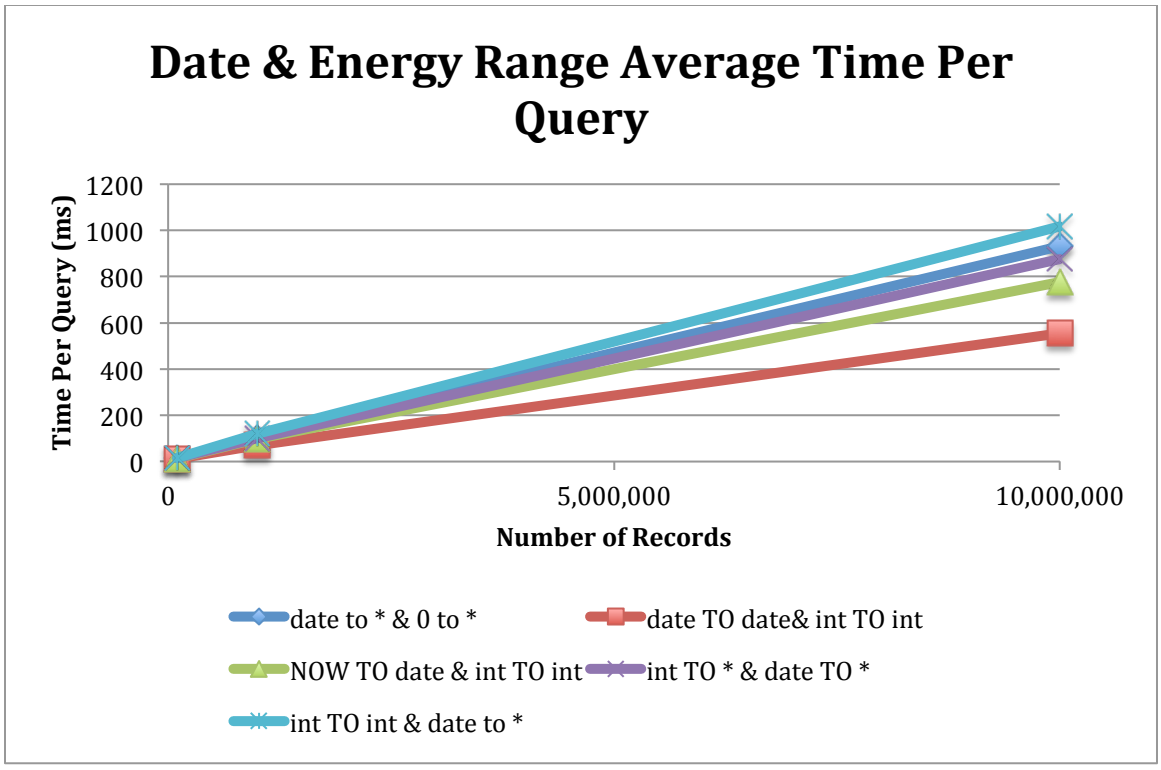
- createdDate:[date TO *] AND finalEnergy:[energy TO *] : 600 queries
- createdDate:[date TO date] AND finalEnergy:[energy TO energy] : 600 queries
- createdDate: [NOW TO date] AND finalEnergy:[energy TO energy]: 600 queries
- finalEnergy: [energy TO *] AND createdDate:[date TO *]: 600 queries
- finalEnergy: [energy TO energy] AND createdDate: [date TO *]: 600 queries

Each of the 5 types of queries were run 600 times with randomly generated dates and energy values.

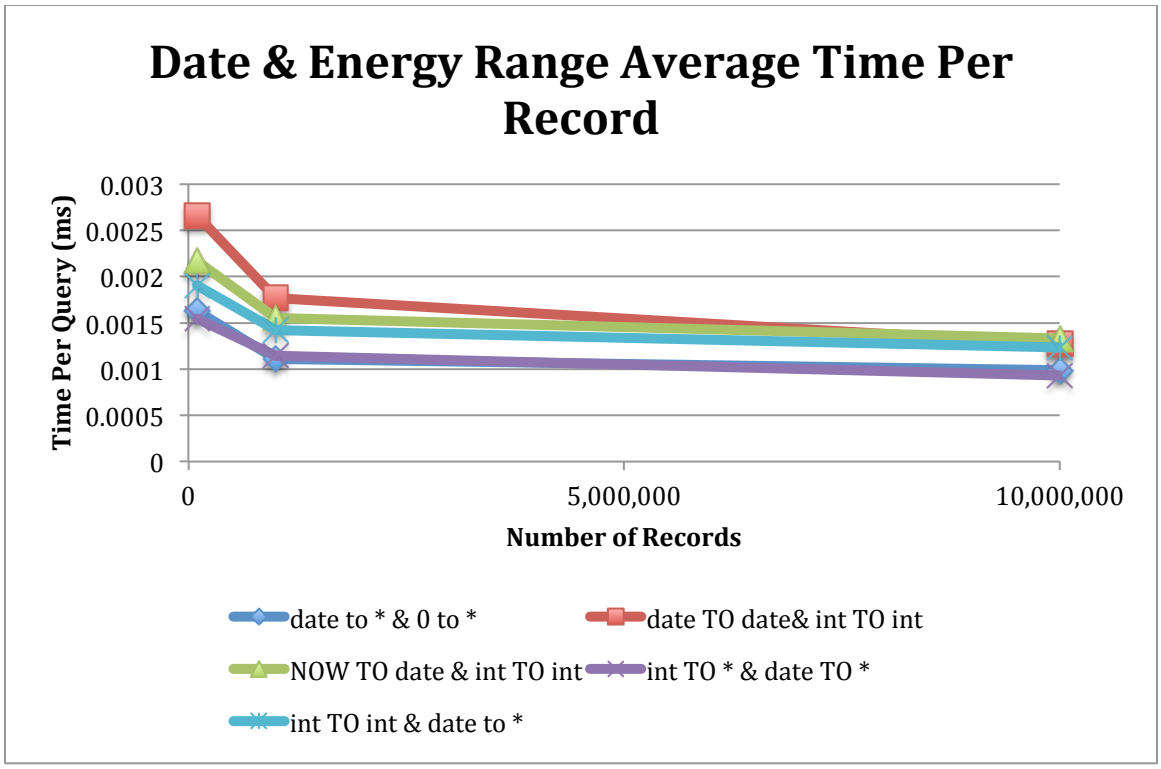
2.3.1 Average time per record (total) vs number of records



2.3.2 Average time per query vs Number of records

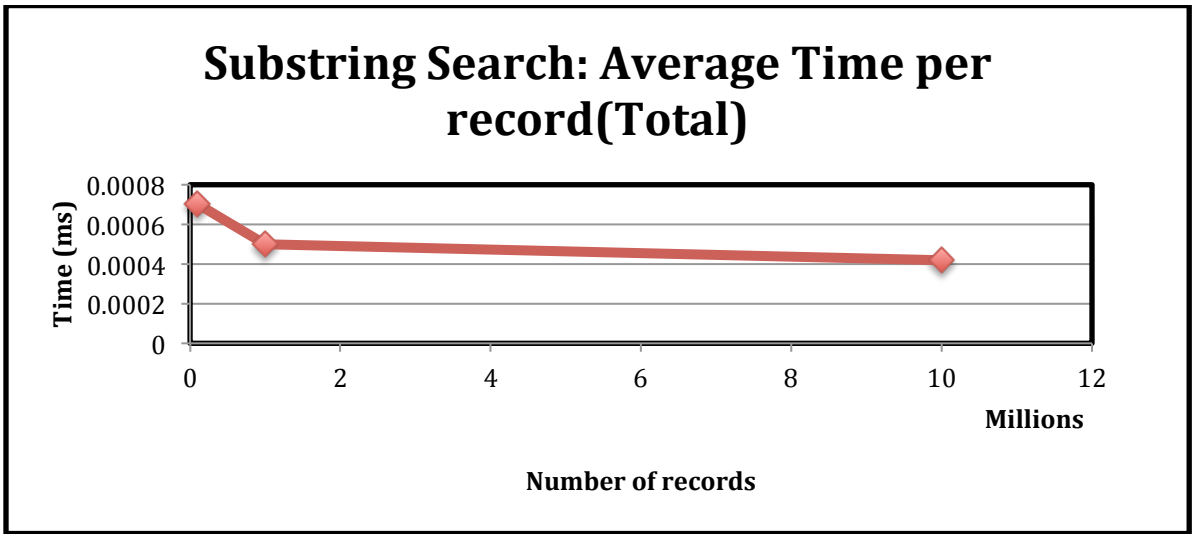


2.3.3 Average time per records (grouped by query type) vs Number of records

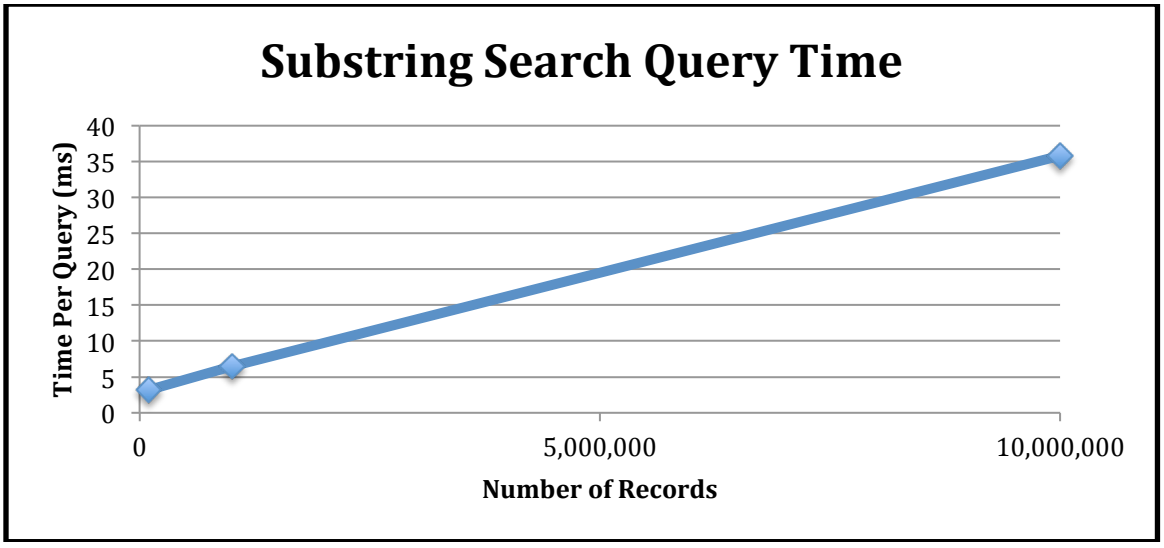


2.4 Substring queries (3000 Queries)

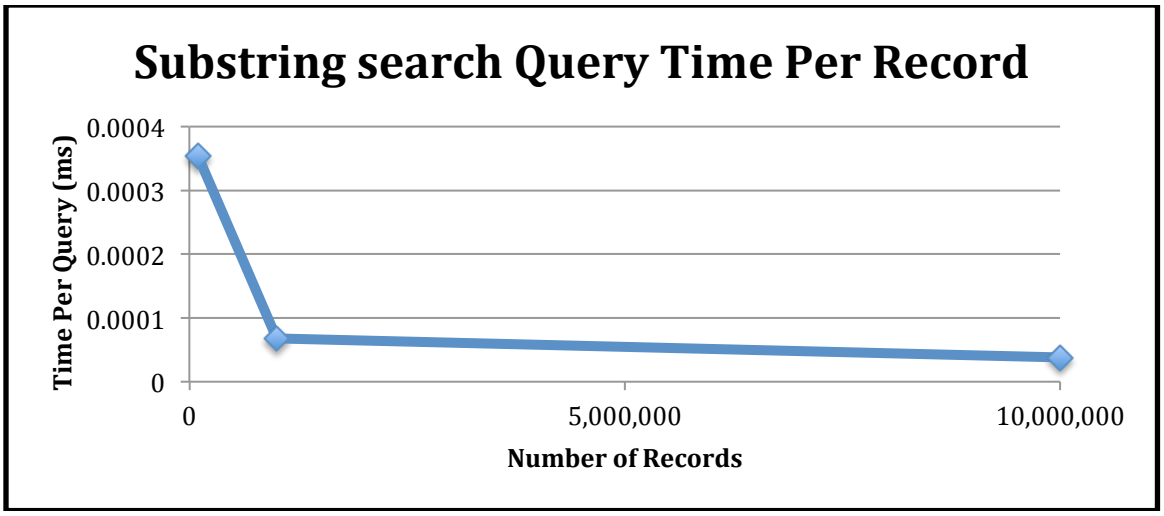
2.4.1 Average time per record (total) vs number of records



2.4.2 Average time per query vs Number of records



2.4.3 Average time per records (grouped by query type) vs Number of records



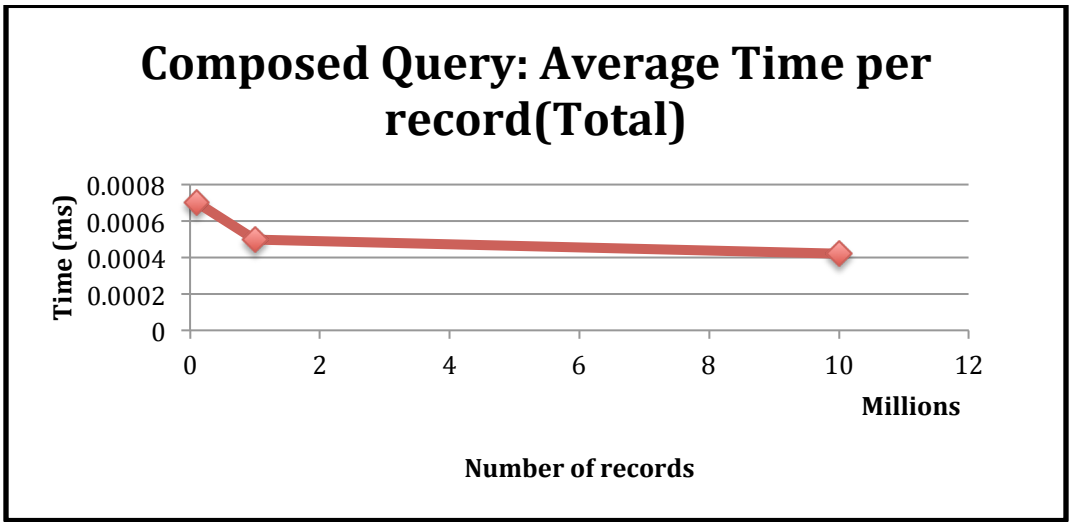
2.5 Substring, Data and Energy Range queries (3000 Queries)

Executed queries:

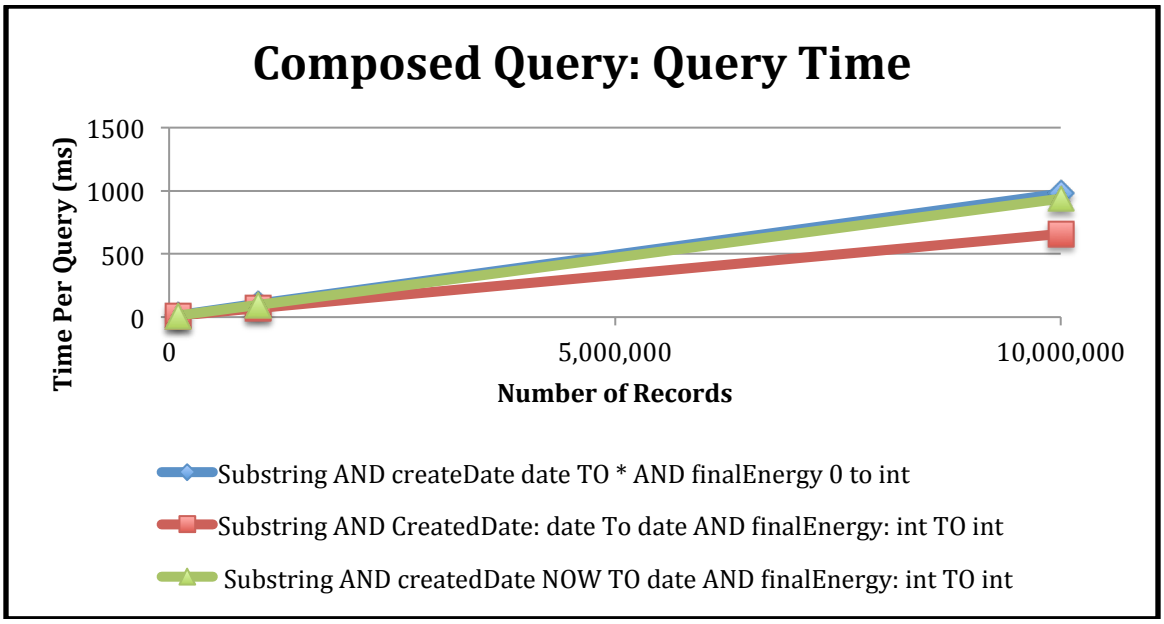
- Substring AND createdDate:[date TO *] AND finalEnergy:[0 TO *] : 1000 queries
- Substring AND createdDate:[date TO date] AND finalEnergy:[energy TO energy] : 1000queries
- Substring AND createdDate: [NOW TO date] AND finalEnergy:[energy TO energy]: 1000queries

Each of the 3 types of queries were run 1000 times with randomly generated dates, energy values and substrings.

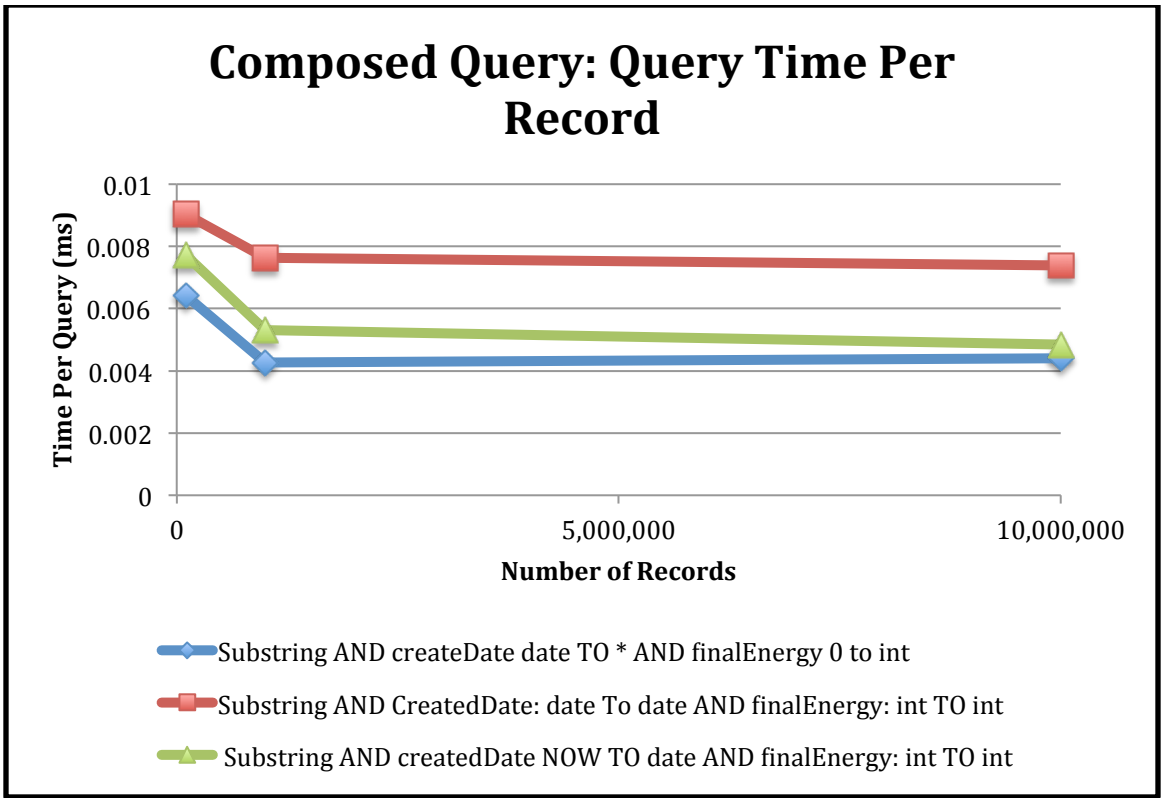
2.5.1 Average time per record (total) vs number of records



2.5.2 Average time per query vs Number of records



2.5.3 Average time per records (grouped by query type) vs Number of records


Composed Query: Query Time Per Record
Time Per Query (ms)
0.01
0.008
0.006
0.004
0.002
0
0
5,000,000
10,000,000
Number of Records
Substring AND createDate date TO * AND finalEnergy 0 to int
Substring AND CreatedDate: date To date AND finalEnergy: int TO int
Substring AND createdDate NOW TO date AND finalEnergy: int TO int