Programming Parallel Algorithms Project

Project Title: Design and Implementation of a Parallel Relational Database from scratch

Project Group Member(s):

1. Darshan Dinesh Kumar (dd3888, N10942768)

Background

- Data is growing exponentially
- Storing and managing this vast data is complex
- Databases are the ubiquitous solutions
- Designing large-scale, efficient databases is challenging
- Advent of Multicore and Multiprocessor systems
- Parallelism as a potential candidate to improve performance

Overview of Implementation

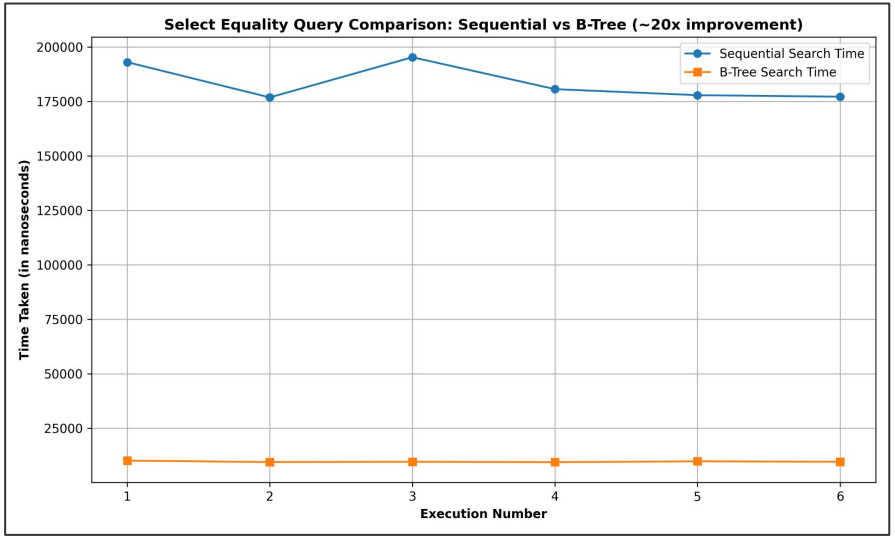
- **Parallelizing the queries**
  - Parallel For Loop (Conditional Selection and Projection)
  - Parallel Merge Sort (Order by and Sorting as a prerequisite)
  - Parallel primitives like tabulate, filter, reduce (Group by and Aggregation)

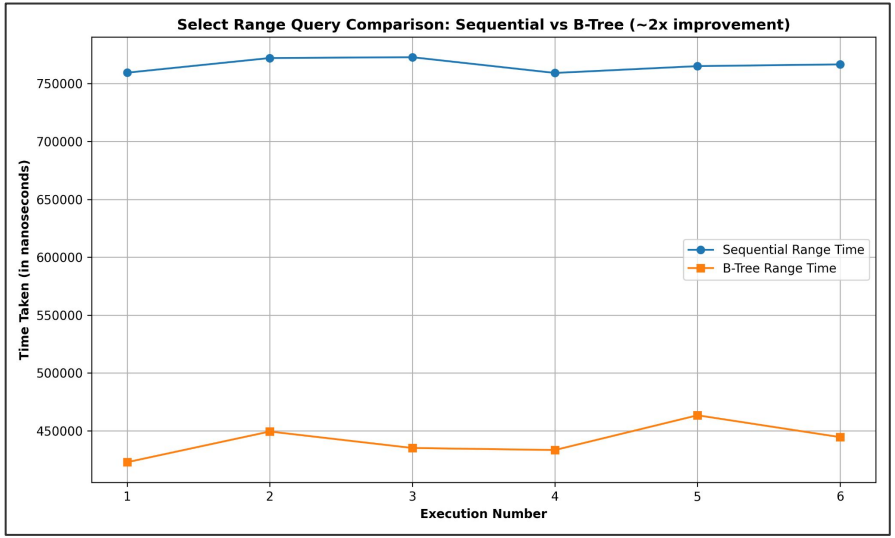
- **B-Tree as an indexing data structure**
  - Search (Conditional selection based on primary key)
  - Range Traversal (Selection within a range of primary key values)

- **Functional Testing**

- **Performance Testing**
  - Person Relation with fields: *id(PK), fname, lname, age, country, salary*
  - 10000 rows with randomly generated data
  - Results generated on *crunchy5.cims.nyu.edu* with 32 Cores each with 2 threads

Results



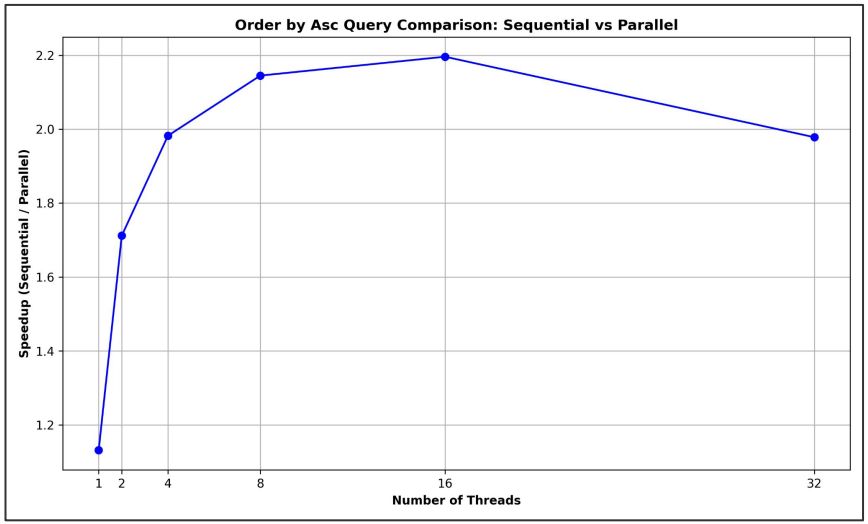


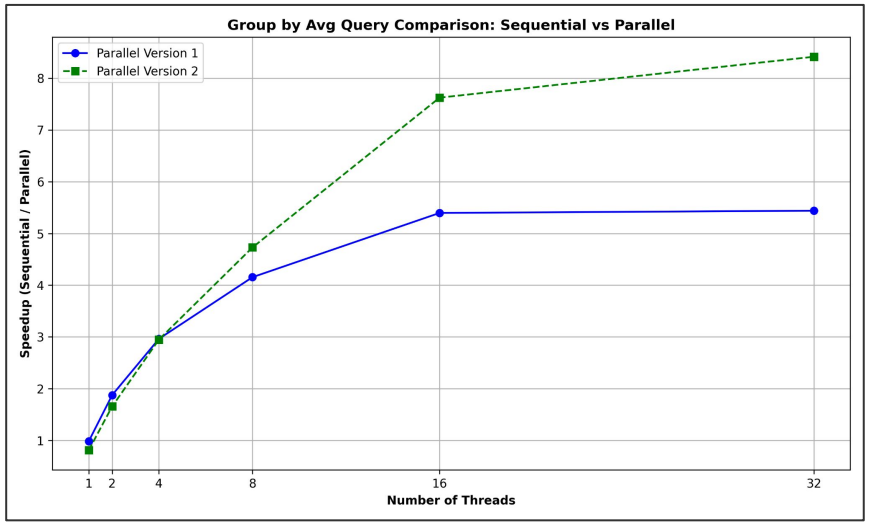

```
SELECT *
FROM Person
WHERE id == 7312
```

```
SELECT *
FROM Person
WHERE id >= 1291 AND id <= 1524
```

NYU

Results






```
SELECT *
FROM Person
ORDER BY fname, lname, id ASC
```

```
SELECT AVG(salary), country
FROM Person
GROUP BY country
```

** **Parallel Version 1:** Parallel Merge Sort followed by sequential traversal
Parallel Version 2: Parallel Merge Sort followed by algorithm using tabulate, filter, reduce

THANK YOU !

NYU