WriteUp for Lab4

521120910234 Shiyi Huang

Implementation

Selectivity Estimation

- **Approach**: Describe the implementation of selectivity estimation for filters using histograms. Mention the choice of using **IntHistogram** for integer fields and **StringHistogram** for string fields, and the rationale behind it.
- **Challenges**: Discuss any challenges faced while implementing histograms, such as determining the number of buckets or handling skewed data.

Cost-Based Optimizer

- **Join and Scan Cost Estimates**: Explain how you calculated the costs for table scans and joins. Detail the formulae used and any assumptions made regarding I/O and CPU costs.
- **Join Ordering**: Describe the implementation of the **orderJoins** method, which is crucial for the optimizer. Explain how you utilized the Selinger optimizer approach to determine the optimal join order based on estimated costs.
- **Optimization Techniques**: Highlight any specific optimization techniques you applied, such as dynamic programming or greedy algorithms, to improve the efficiency of join ordering.

Conclusion

In conclusion, this lab provided a valuable opportunity to extend SimpleDB's capabilities and deepen my understanding of database operations. By implementing operators for table modifications, selections, joins, and aggregates, I gained hands-on experience with core database concepts. The challenges faced during implementation helped solidify my understanding of database internals and provided insights into

optimizing query performance. Overall, this lab was an enriching learning experience that reinforced key concepts in database systems.

Time Spent

I spent approximately 15 hours on this lab assignment. The majority of the time was dedicated to implementation and debugging, with additional time spent on reading documentation and understanding the provided codebase.