

Acknowledgements

I am deeply grateful to Prof. Jayant R. Haritsa for his unmatched guidance. He was and always will be an inspiration for me for his thoughts and his perception of looking at things around us, be it academia or real life. It is my good luck, that I got a chance to work with him and obtain few chunks of knowledge from him.

I am thankful to Anshuman Dutt, Srinivas Karthik and Sanket Purandare for their guidance and discussions. I would express my gratitude towards my lab mates for suggestions during my work. My sincere thanks goes to my friends Diksha and Hemant, who have patiently listened and understood many complications, and help me bring out much clear pictures to work with. Also, CSA and CDS office staff was the team of people who make academic functioning smooth with best co-ordination possible.

Finally, I would express my gratitude to my parents, who have taught me how to stay calm and keep working on what is essential, nothing I have done or will ever do will match their support that I carry within.

Abstract

Declarative query processing in database systems often leads to sub-optimal performance due to wrong selectivity estimation from those encountered during actual execution. Plan Bouquets is a technique proposed to substitute selectivity estimation by selectivity discovery at run-time, to provide worst case performance guarantees. This is done by performing multiple partial executions for same query in an incremental fashion of cost budget from a bouquet which is compiled at very first stage.

This technique is suitable for OLAP queries as high overheads of optimizing most part of selectivity error space are amortized over multiple invocation of query in OLAP scenarios. Full space exploration overheads in past are improved upon with NEXUS, which is an algorithm, that only discovers points useful for bouquet compilation.

In this work, we proposed an adaptive version of NEXUS named AdaNEXUS, which utilize geometrical properties of contours to be discovered for bouquet. This algorithm reduces overheads of compilations empirically keeping the worst case performance complexity same. Further, we provide upper bounds for maximum cost deviation possible during bouquet compilation due to use for either NEXUS or AdaNEXUS. Evaluation of proposed system is done on TPC-DS benchmark with different scales to test system. It is demonstrated that around an order of magnitude reduction is observed in compilation overheads when compared with NEXUS, also quality of plans discovered by AdaNEXUS is better than those discovered by NEXUS.