Universiteit
Leiden
The Netherlands

Progressive Indexes

Timbó Holanda, P.T.

Citation
Timbó Holanda, P. T. (2021, September 21). *Progressive Indexes*. *SIKS Dissertation Series*. Retrieved from https://hdl.handle.net/1887/3212937

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/3212937

Note: To cite this publication please use the final published version (if applicable).

Summary

Interactive exploration of large volumes of data is increasingly common, as data scientists attempt to extract interesting information from large opaque data sets. This scenario presents a difficult challenge for traditional database systems, as (1) nothing is known about the query workload in advance, (2) the query workload is constantly changing, and (3) the system must provide interactive responses to the issued queries. This environment is challenging for index creation, as traditional database indexes require upfront creation, hence a priori workload knowledge, to be efficient.

In this work, we introduce *Progressive Indexing*, a novel performance-driven indexing technique that focuses on automatic index creation while providing interactive response times to incoming queries. Its design allows queries to have a limited budget to spend on index creation. The indexing budget is automatically tuned to each query before query processing. This allows for systems to provide interactive answers to queries during index creation while being robust against various workload patterns and data distributions.

We develop progressive algorithms to index one and multiple dimensions. In addition, we introduce *Progressive Merges*, a robust algorithm that merges appends into our Progressive Indexes without penalizing single queries.