Fast and Accurate Cardinality Estimation of Multi-Join Queries on Streams and Databases

Tech ID: 33599 / UC Case 2024-970-0

CONTACT

Edward Hsieh
hsiehe5@uci.edu
tel: 949-824-8428.

BRIEF DESCRIPTION

Efficiently analyzing large volumes of information, as found in streaming data and big data applications, requires accurate cardinality estimates. This invention is capable of more accurately estimating cardinalities while using little memory and compute, as a result, speeding up query evaluation by as much as 50%.

SUGGESTED USES

·Integration into a query optimizer by database company

·Realtime analysis of financial data for risk exposure and detection of financial crimes

·Oil and gas companies analyzing large volumes of geological data

·Social media companies analyzing user and advertiser interactions

·Industrial applications to analyze sensors throughout a manufacturing process

OTHER INFORMATION

ADVANTAGES

·Faster, more memory efficient, and more accurate cardinality estimation of queries that enables better query optimization and faster evaluation

·The invention requires only a single pass over the data, making it applicable to streaming data

CATEGORIZED AS

» **Computer**
   » Software

FULL DESCRIPTION

Large databases and big data applications are important components of modern digital systems. In order to efficiently evaluate queries with the increasing scale of data, it is critical for query optimizers to determine an appropriate join order. At their core, query optimizers rely on cardinality estimates to make their decisions. This invention enables efficient estimation of cardinalities using little memory, by generating small sketches of the data. The sketches are created in a single pass over the data, in arbitrary order, enabling applications to streaming data. Streaming data naturally arises in many big data applications, including network traffic monitoring, recommendation systems, natural language processing, financial systems, and widespread deployment of industrial sensors. The invention can estimate the cardinality of arbitrary multi-join queries, which allows for better optimization of complex data analysis from multiple sources.Also it is orders of magnitude faster than other cardinality estimators, and more accurate, resulting in as much as 50% faster evaluation of queries.Using this invention can reduce the amount of computing power and memory needed to perform complex analysis of large real-time data sets.

RELATED CASES

2024-970-0

STATE OF DEVELOPMENT

Working implementation tested on real data sets and evaluated in PostgreSQL. No end-to-end integration with existing data management systems.

PATENT STATUS

Country	Type	Number	Dated	Case
United States Of America	Published Application	20250258817	08/14/2025	2024-970

RELATED MATERIALS

» Convolution and Cross-Correlation of Count Sketches Enables Fast Cardinality Estimation of Multi-Join Queries. Mike Heddes, Igor Nunes, Tony Givargis, Alex Nicolau - 05/14/2024