Data Stream Management

Synthesis Lectures on Data Management

Editor

M. Tamer Özsu, ***University of Waterloo***

Synthesis Lectures on Data Management is edited by Tamer Özsu of the University of Waterloo. The series will publish 50- to 125 page publications on topics pertaining to data management. The scope will largely follow the purview of premier information and computer science conferences, such as ACM SIGMOD, VLDB, ICDE, PODS, ICDT, and ACM KDD. Potential topics include, but not are limited to: query languages, database system architectures, transaction management, data warehousing, XML and databases, data stream systems, wide scale data distribution, multimedia data management, data mining, and related subjects.

Data Stream Management
Lukasz Golab and M. Tamer Özsu
2010

Access Control in Data Management Systems
Elena Ferrari
2010

An Introduction to Duplicate Detection
Felix Naumann and Melanie Herschel
2010

Privacy-Preserving Data Publishing: An Overview
Raymond Chi-Wing Wong and Ada Wai-Chee Fu
2010

Keyword Search in Databases
Jeffrey Xu Yu, Lu Qin, and Lijun Chang
2009

Data Stream Management

Lukasz Golab and M. Tamer Özsu

ISBN: 978-3-031-00709-5 paperback
ISBN: 978-3-031-01837-4 ebook

DOI 10.1007/978-3-031-01837-4

A Publication in the Springer series
SYNTHESIS LECTURES ON DATA MANAGEMENT

Lecture #5
Series Editor: M. Tamer Özsu, *University of Waterloo*
Series ISSN
Synthesis Lectures on Data Management
Print 2153-5418 Electronic 2153-5426

Data Stream Management

Lukasz Golab
AT&T Labs—Research, USA

M. Tamer Özsu
University of Waterloo, Canada

SYNTHESIS LECTURES ON DATA MANAGEMENT #5

ABSTRACT

In this lecture many applications process high volumes of streaming data, among them Internet traffic analysis, financial tickers, and transaction log mining. In general, a data stream is an unbounded data set that is produced incrementally over time, rather than being available in full before its processing begins. In this lecture, we give an overview of recent research in stream processing, ranging from answering simple queries on high-speed streams to loading real-time data feeds into a streaming warehouse for off-line analysis.

We will discuss two types of systems for end-to-end stream processing: Data Stream Management Systems (DSMSs) and Streaming Data Warehouses (SDWs). A traditional database management system typically processes a stream of ad-hoc queries over relatively static data. In contrast, a DSMS evaluates static (long-running) queries on streaming data, making a single pass over the data and using limited working memory. In the first part of this lecture, we will discuss research problems in DSMSs, such as continuous query languages, non-blocking query operators that continually react to new data, and continuous query optimization. The second part covers SDWs, which combine the real-time response of a DSMS by loading new data as soon as they arrive with a data warehouse's ability to manage Terabytes of historical data on secondary storage.

KEYWORDS

Data stream Management Systems, Stream Processing, Continuous Queries, Streaming Data Warehouses

Contents