# Chapter 3.11
# View Management Techniques and Their Application to Data Stream Management

**Christoph Quix**
*RWTH Aachen University, Germany*

**Xiang Li**
*RWTH Aachen University, Germany*

**David Kensche**
*RWTH Aachen University, Germany*

**Sandra Geisler**
*RWTH Aachen University, Germany*

## ABSTRACT

Data streams are continuous, rapid, time-varying, and transient streams of data and provide new opportunities for analysis of timely information. Data processing in data streams faces similar challenges as view management in data warehousing: continuous query processing is related to view maintenance in data warehousing, multi-query optimization for continuous queries is highly related to view selection in conventional relational DBMS and data warehouses. In this chapter, we give an overview of view maintenance and view selection methods, explain the fundamental issues of data stream management, and discuss how view management techniques from data warehousing are related to data stream management. We also give directions for future research in view management, data streams, and data warehousing.

## INTRODUCTION

The management of views is a fundamental problem in the design and maintenance of data warehouse systems. Materialized views speed up query processing, but require additional storage and need to be maintained in case of updates of the base data. In order to balance the efficiency of query processing and view maintenance, *view selection* techniques

DOI: 10.4018/978-1-60566-816-1.ch005

have been proposed which select a set of views that approximates optimal costs for query processing and view maintenance.

Data warehouses rely heavily on analysis of up-to-date information to support decision makers. The advent of a new class of data management applications, namely data stream management systems (DSMS), provides new opportunities for analysis of timely information. A data stream is a continuous, rapid, time-varying, and transient stream of data. There are connections between DSMS and view management. Whereas continuous query processing is related to view maintenance in data warehousing, multi-query optimization for continuous queries is highly related to view selection in conventional relational DBMS and data warehouses. In this chapter, we give an overview of view maintenance and view selection methods, explain the fundamental issues of data stream management, and discuss how view management techniques from data warehousing are related to data stream management.

The chapter is structured as follows: section 2 briefly explains the roles of views in data warehouses. Section 3 gives an overview of view maintenance methods and classifies them according to various criteria. Then, section 4 explains the view selection problem and presents a taxonomy of existing view selection techniques. Section 5 discusses issues and challenges in data stream management and summarizes recent results in research on data streams. Section 6 discusses the relationship of view management techniques to data stream management. Similarities, differences and possible connections between data stream management and view management are discussed. Finally, section 7 summarizes the chapter and points out directions for future research in view management, data streams, and data warehousing.

## Views in Data Warehousing

A view can select or restructure data in such a way that an application can use the data more efficiently. Different from On-Line Transaction Processing (OLTP) systems, which focus at managing the common data operations, data warehouses aim at supporting data analysis (i.e., On-Line Analytical Processing, OLAP) and are known for their vast volume of data and complexity of queries. The response time of queries, if evaluated from base tables, is usually too long for users to tolerate as a huge amount of data has to be processed. Therefore, it is a common practice to pre-compute summaries of base tables in order to reduce the query response time. The following example illustrates the benefit of materializing views:

**Example 1** Consider the TPC-D benchmark (Serlin, 1993), modeling a data cube of sales with three dimensions: part, supplier, and customer. We denote the base table as *R(part; supp; cust; sales)*. The following query is posed by users:

*Q*: SELECT part, SUM(sales) AS total
FROM R
GROUP BY part;

The following two materialized views can both benefit Q:

$V_1$: SELECT part, cust, SUM(sales) AS total
FROM R
GROUP BY part, cust;

$V_2$: SELECT part, supp, SUM(sales) AS total
FROM R
GROUP BY part, supp;

It depends on the statistics of the data to decide which view is better in terms of query response or storage cost. For instance, the statistics of the TPC-D database are as follows:

- $R$: 6M rows
- $V_1$: 6M rows
- $V_2$: 0.8M rows