Lesson 2

Data Stream Management System (DSMS)

Streaming Data

- Requires parsing and processing at once when they arrive in the system.

- Responses against the queries are precise in static DBMS

- Streaming data can unpredictably change in both size and frequency, and thus results into approximate responses

"Big Data Analytics ", Ch.07 L02: Data Stream Mining ...Spark Streaming, Raj Kamal and Preeti Saxena, © McGraw-Hill Higher Edu. India

Two types of data streams

(1) Transactional data stream: They carry data related to the interactions between different entities, for example:

(i) ATM transactions

(ii) Telecommunication

(iii) Web access

Two types of data streams

(1) Transactional data stream: They carry data related to the interactions between different entities, for example:

(i) ATM transactions

(ii) Telecommunication

(iii) Web access

Two types of data streams

2. Measurement data stream: They carry data related to measured values or metric of entity states. For example:

(i) Sensor networks

(ii) IP network

(iii) Climate data

Data Stream Management System (DSMS)

- Refers to a data-intensive application in which the data models as a transient data stream

- An application program that manages streaming data

- The usage of the program─ Distinct from using of persistent relations in DBMS.

DSMS

- Provides on-line analysis of rapidly changing stream of data

- Suitable for real-time, continuous, ordered (arrival time or timestamp) sequence of data elements. Also, for data that is large to store entirely and continuous querying environment

DSMS

- Transient stream (on-line analysis)

- The extraction generally depends on the specified threshold

- Continuous queries

- Implements sequential access

- Bounded main memory

Commercial databases for streams

(i) StreamBase (commercial version of Aurora/Borealis)

Commercial databases for streams

(ii) Truviso (commercial version of TelegraphCQ); Cisco acquired it in May 2012 extending existing relational databases to build TelegraphCQ and Truviso syntaxes. They are extensions of PostgreSQL that incorporate data stream

Commercial databases for streams

(iii) TIBCO Business Events, Oracle Business Activity Monitoring

Summary

We learnt:

- DSMS

- Characteristics of DSMS

- Provides on-line analysis of rapidly changing stream of data

- Suitable for real-time, continuous, ordered (arrival time or timestamp) sequence of data elements

End of Lesson 2 on
Data Stream Management System