

Lesson 1

Data Stream Concepts and Models

Stream

- Refers to a sequence of data elements or symbols made available over time
- Data stream transmits from a source and receives at the processing end in a network
- A continuous stream of data flows between the source and receiver ends, and which is processed in real time

Stream

- Also refers to communication of bytes or characters over sockets in a computer network
- A program uses stream as an underlying data type in inter-process communication channels.

Examples of Data Stream Applications

1. Making data-driven marketing decisions in real time. It requires the use of data from trends analyses of real-time sales, and analysis of social media, and the sales distribution.
2. Monitoring and detection of potential failures of system using network management tool

Examples of Data Stream Applications

3. Monitoring of industrial or manufacturing machinery in real time
4. A sensor network or IoT controlled by another entity, or a set of entities
5. Watching online video lectures, and rewinding or forwarding them

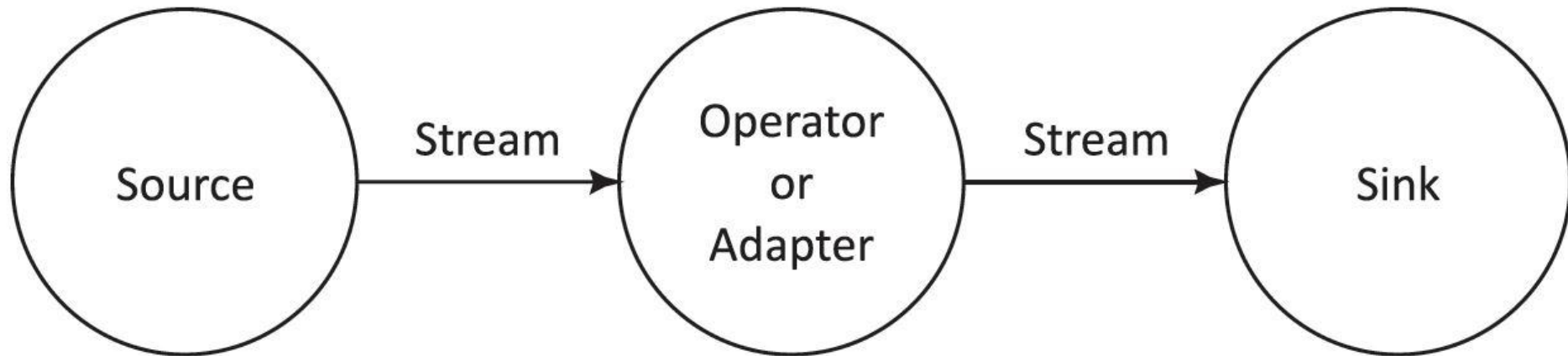
Application processing of a data stream

- Processing is in micro-batches instead of processing batches
- Processing of stream can be comprehended as filling milk in bottles on a conveyor belt and capping them, one at a time successfully rather than in a large batch at the same time

Data Stream Model

- Stream is data in motion
- Three approaches for updating the endpoints (sinks) are (i) non-overlapping, (ii) slow (batch processing) and (iii) fast (near real-time)
- Different ways of modeling data stream, querying, processing and management.

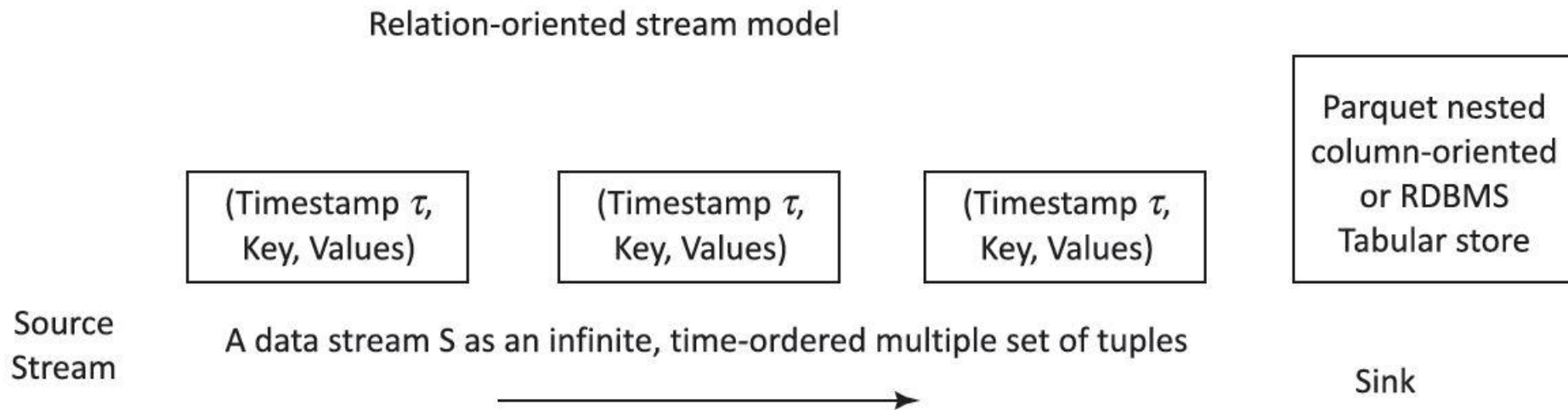
Figure 7.1 Graph-based stream data model for processing at an operator or adapter



Relation-oriented stream-tuples model

- Stream dataflow can be modeled as tuples flow
- Individual data items may be relational tuples in a data stream model

Figure 7.2 Relation-oriented stream-tuples model (Time stamp for real-time streaming data)



Data Stream

- An unbounded and time-ordered sequence of data items (relational tuples) in the data stream model
- The receiving software receives the sequences in order and sees the data items only once. Each tuple consists of a set of attributes, like a row in a database table. The tuples have a schema-like

Data Stream

- Each tuple consists of a set of attributes, like a row in a database table
- The tuples have a schema-like
- traditional database.
- One of the attributes in the tuple schema is a timestamp, usually represented by t .

Data Source and Sink

- Data Source may be a RDBMS row-oriented storage tables or Parquet tables
- Parquet nested column-oriented data stream transfer data to Data sink [Parquet is nested hierarchical columnar-storage concept. Nesting sequence is table, row groups, column chunk, page chunks with the column chunk.]

Object-based data stream model

- Data-flows modeled as objects
- Examples: Cougar and Tribeca object-based data stream
- Cougar models sensors' data as a stream of objects
- Tribeca models the network monitoring data as a stream of objects

XML-based data stream model

- Example: NiagaraCQ, an XML-based data stream model
- Scalable continuous query processing over XML documents
- Performs operations over millions of simultaneous queries by dynamically grouping them according to their structural similarities.

Window-based data stream model

- Stream data direction can be towards fixed window, sliding window or landmark window-sinks (end-points) [Window means a time window during which the data stream is looked at an instance.]

Summary

We learnt:

- Data stream a continuous stream of data flows between the source and receiver ends, and which is processed in real time
- Graph Model of Data Stream
- Relational tuples model

Summary

We learnt:

- Objects based Data Stream transfer from source to sink
- XML based Data Stream
- Time-Windows based data stream

End of Lesson 1 on
**Data Stream Concepts and
Models**