Data Stream Management

Group I
SE-110 Software Engineering (Spring 2012)

Sumit Jain
Uddipan Dutta
Vikas Korjani

IIIT-Bangalore

Feb 16, 2012

Introduction...

- **Data Stream**
  Sequence of data packets flowing across network.
  e.g. traffic at a router in an IP network, road traffic in sensor network, transactional data streams.

- **Data Stream Management System**
  Allows filtering/querying data over streams for various analysis purpose.

What is the need for data stream processing...

- What is wrong with database systems?
- What are some problems for which a traditional DBMS is an awkward fit?

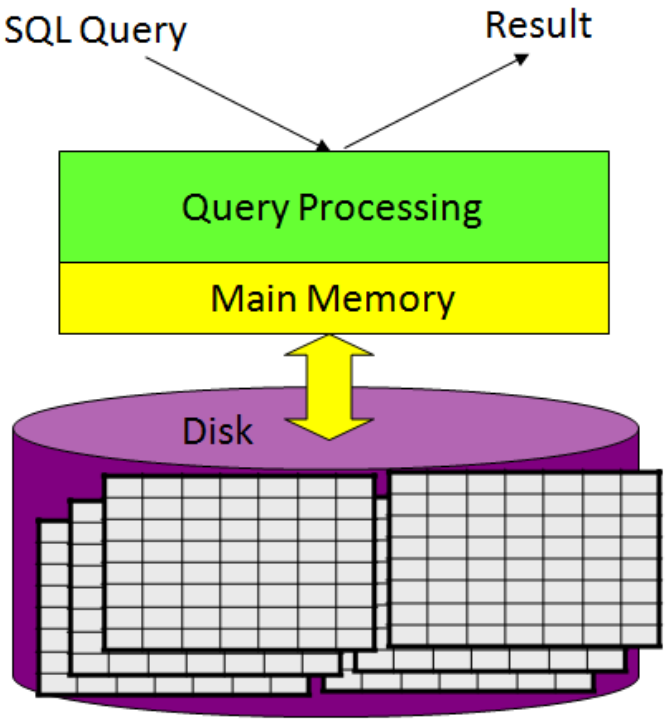
Where are DSMS applications required...

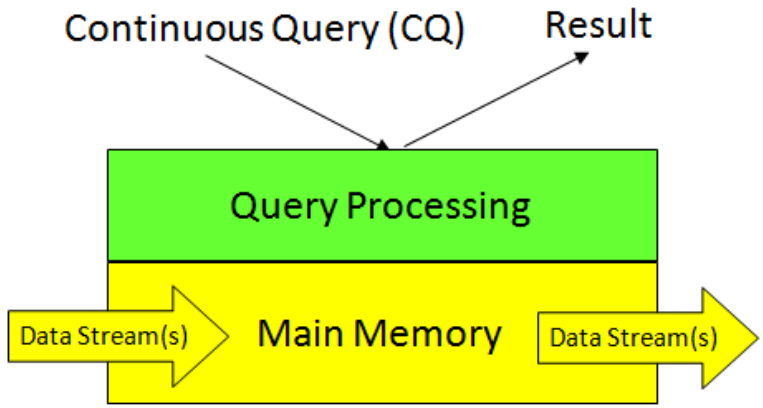
1. **Sensor Networks**
   e.g. TinyDB
2. **Network Traffic Analysis**
   e.g. Traffic statistics and critical condition detection.
3. **Financial Tickers**
   e.g. On-line analysis of stock prices, discover correlations, identify trends.
4. **Transaction Log Analysis**
   e.g. Web click streams and telephone calls.

DSMS

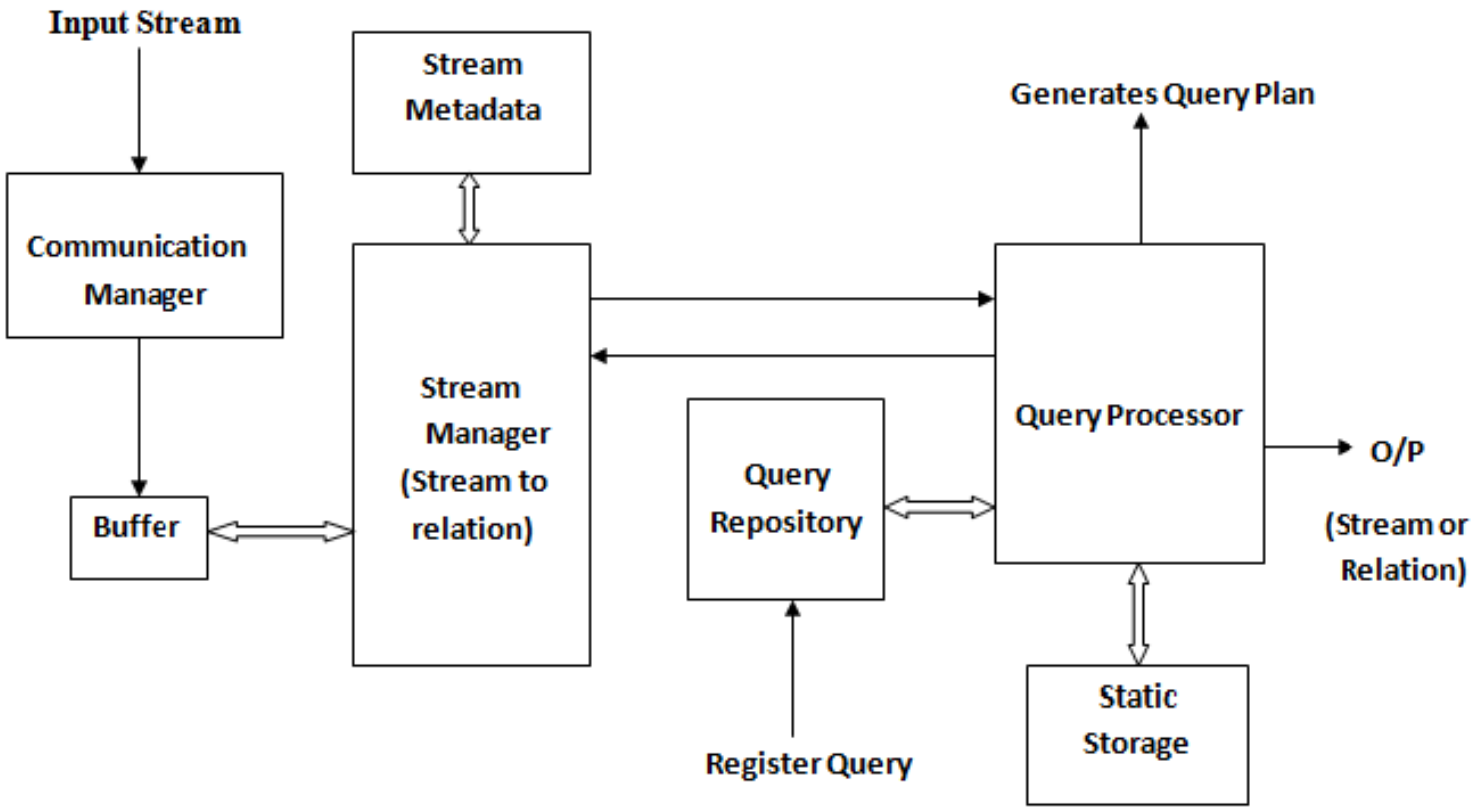
1. Transient streams (on-line analysis)
2. Continuous queries (CQs)
3. Sequential access
4. Bounded main memory
5. Real-time requirements

DBMS vs DSMS...


SQL Query
Result
Query Processing
Main Memory
Disk


Continuous Query (CQ)
Result
Query Processing
Data Stream(s)
Main Memory
Data Stream(s)

Architecture...



Data Stream Management System Architecture

Components...

- **Communication Manager**
  Establishes/configures connection over network and provides subscription to multiple streams of data.
- **Input Buffer**
  To prevent streams to overwhelm our system, we use Input Buffer.
- **Stream Metadata**
  Stream to relation mapping requires prior information about the stream structure which is stored in Stream Metadata.

Components (contd.)...

- **Stream Manager**
  Key responsibility of Stream Manager is stream to relation mapping using information in Stream Metadata.
- **Static Storage**
  Static Storage is used for storing the intermediate states and accessing them.
- **Query Repository**
  Stores all the registered queries on different streams.
- **Query Processor**
  CQL compiles to query plan, has query operators, inter-operator queues, synopses.

Requirements...

1. Manage network connections and configure streams.
2. User should be capable of subscribing to streams provided source details.
3. System should be capable of registering continuous queries on any registered stream with the server.
4. System contains Query Processor that generates execution plans for registered queries.
5. Continuous queries produce an explicit output that can be viewed by user or can be used as sub-queries whose output is used by other queries.

Modules...

- Client interface module
- Server/Network-manager module
- Stream Manager (Stream to relational mapper)
- Query Processor module

Development Plan...

Stage 1 : Requirement gathering and Designing architecture

Stage 2 : Server module

Stage 3 : Stream Manager module

Stage 4 : Query Processor module

Stage 5 : Client module

Stage 6 : Integration and Testing

Stage 7 : Reviewing

Milestones...

Stage 1 : 20-02-2012

Stage 2 : 07-03-2012

Stage 3 : 15-03-2012

Stage 4 : 29-03-2012

Stage 5 : 10-04-2012

Stage 6 : 15-04-2012

Stage 7 : 20-04-2012

Individual Contributions...

Sumit Jain

- Explored Stanford Stream Manager, design documents and implementation plans.
- Contributed in architectural design of DSMS (Data Stream Management System).
- Currently, working on implementation plan.

Individual Contributions (Contd.)...

Uddipan Dutta

- Explored how Stream Data Management can be used for network monitoring.
- POC on capturing network packet from network interface devices and generating stream for DSMS, which can then be analyzed for network monitoring.
- Currently, working on implementation plan.

Individual Contributions (Contd.)...

Vikas Korjani

- Read reference material online on Stream Data Management.
- Contributed in architectural design of DSMS.
- POC, configured Derby DB to create in-memory Database and transient table.
- Currently working on implementation plan.

References...

- Arasu, S. Babu, and J. Widom.
  The cql continuous query language: Semantic foundations and query execution. Technical report, Stanford University Database Group, Oct. 2003.
  Available at: http://dbpubs.stanford.edu/pub/2003-67.

- Minos Garofalakis, Johannes Gehrke, and Rajeev Rastogi. 2007. Data Stream Management: Processing High-Speed Data Streams (Data-Centric Systems and Applications). Springer-Verlag New York, Inc., Secaucus, NJ, USA

Thank You