Module 4

Q. Difference between DSMS and DBMS

=>

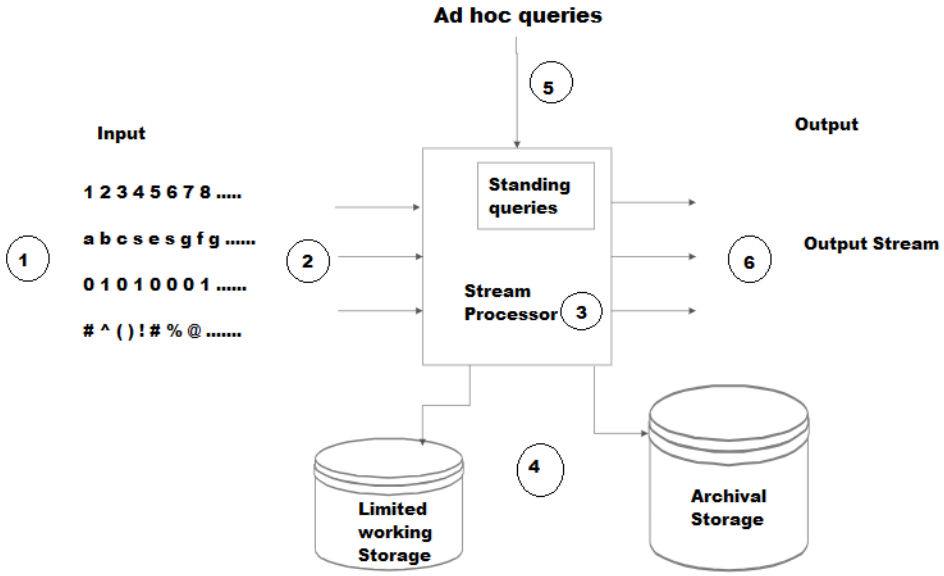
Aspect	DBMS	DSMS
Data Nature	Static, persistent data stored on disk	Dynamic, continuous streams of data
Data Model	Finite, stored relations (tables)	Infinite, time-varying data streams
Queries	One-time queries on stored data	Continuous queries over streaming data
Query Execution	Pull-based (query executes when requested)	Push-based (data pushed to queries continuously)
Response Time	Can be slower, optimized for accuracy	Real-time or near real-time processing
Data Access	Random access to entire dataset	Sequential access, data processed as it arrives
Storage	Data permanently stored in database	Limited storage, data may be discarded after processing
Updates	Explicit INSERT/UPDATE/DELETE operations	Implicit updates as new data arrives
Processing Model	Store-then-process	Process-on-arrival (may store summaries)
Examples	MySQL, PostgreSQL, Oracle, MongoDB	Apache Flink, Apache Storm, StreamBase, Kafka Streams
Use Cases	Banking systems, inventory management, ERP	Stock trading, sensor monitoring, network traffic analysis, social media feeds
Query Language	SQL (declarative, finite queries)	Extended SQL or specialized languages (continuous queries)
Data Volume	Handles large but bounded datasets	Handles unbounded, continuous data flows

Q. Data Stream Management System (DSMS) Architecture

=>

1. A Database Management System (DBMS) is a software system that is designed to manage and organize data in a structured manner.
2. It allows users to create, modify, and query a database, as well as manage the security and access controls for that database.
3. The Data Stream Management System manages continuous streams of data with very fast changes in real-time.
4. Unlike other databases which may be static data, its source might include sensors or social media.

5. It is generally used to deal data streams from with various sources which include sensor data, social media fields, financial reports, etc.
6. Just like DBMS, DSMS also provides a wide range of operations like storage, processing, analyzing, integration also helps to generate the visualization and report only used for data streams.
7. There are wide range of DSMS applications available in the market among them Apache Flint, Apache Kafka, Apache Storm, Amazon kinesis, etc.



DSMS Architecture

DSMS consists of various layer which are dedicated to perform particular operation which are as follows:

1. **Data source Layer**
   i. The first layer of DSMS is data source layer as its name suggest it is comprises of all the data sources which includes sensors, social media feeds, financial market, stock markets etc.
   ii. In the layer capturing and parsing of data stream happens.
   iii. Basically, it is the collection layer which collects the data.
2. **Data Ingestion Layer**
   i. This layer is bridge between data source layer and processing layer.
   ii. The main purpose of this layer is to handle the flow of data i.e., data flow control, data buffering and data routing.
3. **Processing Layer**

i. This layer considers as heart of DSMS architecture it is functional layer of DSMS applications.
ii. It processes the data streams in real time.
iii. To perform processing it is uses processing engines like Apache flink or Apache storm etc.,
iv. The main function of this layer is to filter, transform, aggregate and enriching the data stream.
v. This can be done by derive insights and detect patterns.

4. **Storage Layer**
   i. Once data is processed, we need to store the processed data in any storage unit.
   ii. Storage layer consist of various storage like NoSQL database, distributed database etc.,
   iii. It helps to ensure data durability and availability of data in case of system failure.
5. **Querying Layer**
   i. This layer provides the tools which can be used for querying and analyzing the stored data stream.
   ii. It also has SQL like query languages or programming API.
6. **Visualization and Reporting Layer**
   i. This layer provides tools for perform visualization like charts, pie chart, histogram etc.,
   ii. Based on this visual representation it also helps to generate the report for analysis.
7. **Integration Layer**
   i. This layer responsible for integrating DSMS application with traditional system, business intelligence tools, data warehouses, ML application, NLP applications.
   ii. It helps to improve already present running applications.

For the following algorithms, I would recommend watching the videos, instead of directly going for theory :

1. Bloom Filter : https://youtu.be/s4CORWP3-Ac?si=70cOWSAQ6v1200fN
2. Flajolet Martin : https://youtu.be/SP6yMOGLfPc?si=W5w59vf3IddEjskR
3. Datar Gionis Indyk Motwani (DGIM) : https://youtu.be/uFKWc2YR5MU?si=JfBN_GR_CanioYFG