EXPERIMENT NO: 06

Aim: To implement Flajolet Martin Algorithm using Python programming language.

System software requirements: Python

Theory:

* **Data Stream Management System (DSMS)**

- Any number of streams can enter the system. Each stream can provide elements at its own schedule; they do not have the same data rates or data types, and the time between elements of one stream need not be uniform.
- The fact that the rate of arrival of stream elements is not under the control of the system distinguishes stream processing from processing of data that goes on within a database-management system.
- The latter system controls the rate at which data is read from the disk, and therefore never has to worry about data getting lost as it attempts to execute queries.
- Streams may be archived in a large archival store, but we assume it is not possible to answer queries from the archival store
- A data stream management system (DSMS) is a computer software system to management of data streams. It is similar to DBMS.

* **Count-Distinct problem**

- The count-distinct problem is the problem of finding the number of distinct elements in a data stream with repeated elements.

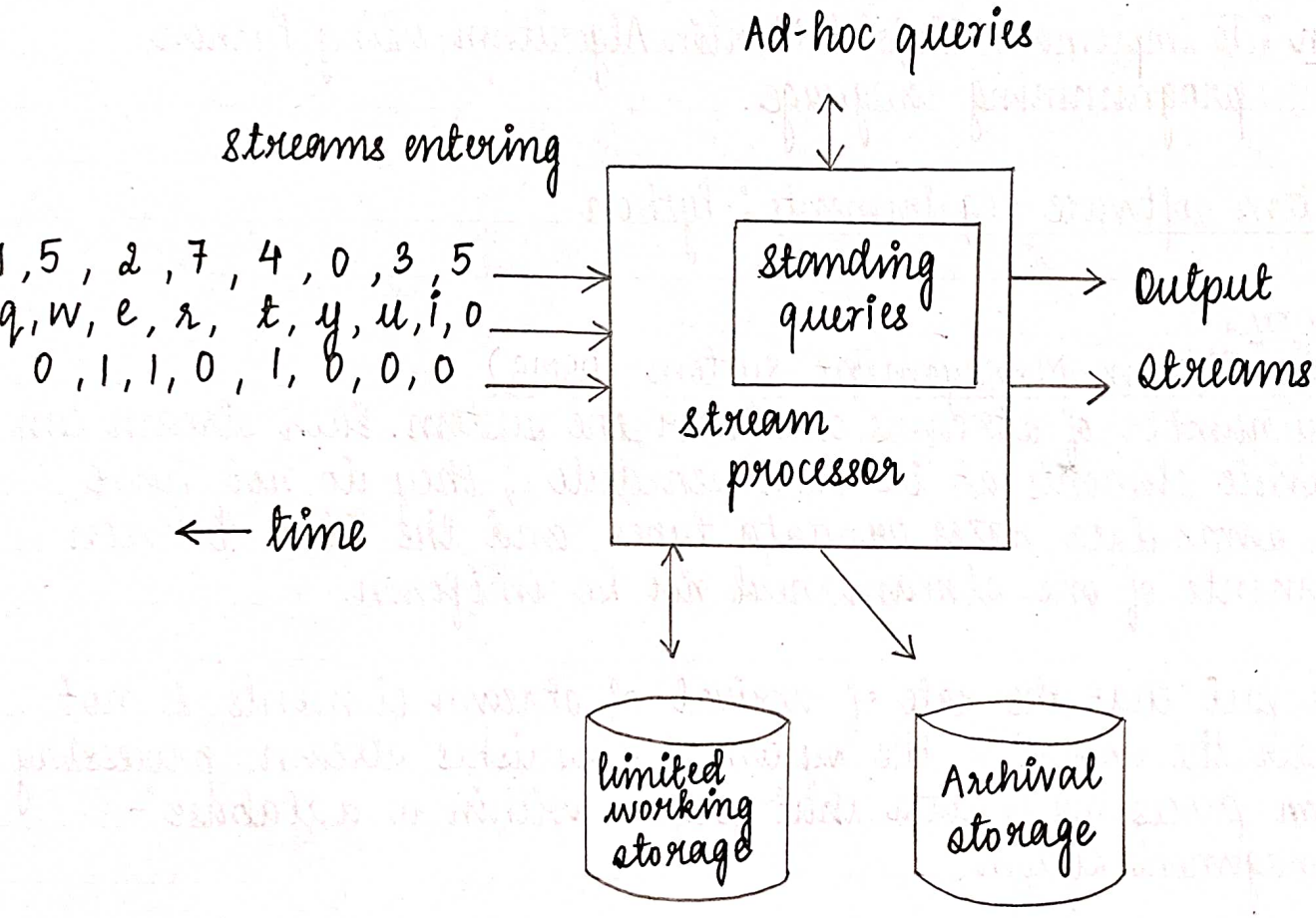

Fig: Data-stream management system

- This is a well-known problem with numerous applications.

- The elements might represent IP addresses of packets passing through a router, unique visitors to web site, elements in a large database, motifs in a DNA sequence, or elements of RFID/ sensor networks.

* Flajolet Martin Algorithm :
- FM algorithm is used to approximate the number of unique elements in a data stream or database in one pass.

- It uses less memory space while executing.

- Pseudo-code :
  1. Selecting a hash function h so each element in the set is mapped to a string to atleast $\log_2 n$ bits.
  2. For each element x, $r(x)$ = length of trailing zeroes in $h(x)$
  3. $R = \max(r(x)) \Rightarrow$ distinct elements $= 2^R$

Conclusion :
Thus we have successfully implemented Flajolet Martin Algorithm using Python language.